

BBVA

FRONTERAS DEL CONOCIMIENTO



FRONTERAS DEL CONOCIMIENTO

FRONTERAS DEL CONOCIMIENTO

JANET ABBATE

SERGIO ALONSO

JESÚS AVILA

ABHIJIT V. BANERJEE

FRANCISCO CALVO SERRALLER

PAUL E. CERUZZI

CARLOS M. DUARTE

JOAN ESTEBAN

LUIS FERNÁNDEZ-GALIANO

JOHN B. HEYWOOD

GERALD HOLTON

ALEXANDER KIND

CAYETANO LÓPEZ

JOAN MASSAGUÉ

JOSÉ M. MATO

ROBERT MCGINN

GINÉS MORATA

LUIS DE PABLO

NATHAN ROSENBERG

VICENTE SALAS FUMÁS

FRANCISCO SÁNCHEZ MARTÍNEZ

JOSÉ MANUEL SÁNCHEZ RON

ANGELIKA SCHNIEKE

SANDIP TIWARI

BBVA

Este libro se concibe como una contribución de BBVA a una tarea extraordinariamente ambiciosa: poner al alcance del lector no especializado una visión rigurosa del estado del arte y las perspectivas de las ramas del conocimiento más características de nuestra época.

Prestigiosos investigadores de todo el mundo, que trabajan en la «frontera del conocimiento», resumen lo más esencial de lo que hoy conocemos, y de lo que aspiramos a conocer en el futuro próximo en los campos de la física, la biomedicina, las tecnologías de la información y las telecomunicaciones, la ecología y el cambio climático, la economía, la empresa y el desarrollo, y analizan el papel de la ciencia y de las artes en nuestra sociedad y en nuestra cultura.

FRANCISCO GONZÁLEZ	Ciencia, innovación y sociedad: desplazando la frontera de lo posible	11
	CIENCIAS BÁSICAS	27
GERALD HOLTON	El lugar de la ciencia en nuestra cultura en el fin de la «era moderna»	31
FRANCISCO SÁNCHEZ MARTÍNEZ	Estructura y evolución del Universo	51
JOSÉ MANUEL SÁNCHEZ RON	El mundo después de la revolución: la física en la segunda mitad del siglo xx	63
SANDIP TIWARI Y ROBERT MCGINN	El arte de lo invisible. Logros, beneficios sociales y desafíos de la nanotecnología	93
	TECNOLOGÍAS DE LA INFORMACIÓN Y LA TELECOMUNICACIÓN	105
PAUL E. CERUZZI	Historia de la informática	109
PAUL E. CERUZZI	Los ordenadores y la exploración espacial	129
JANET ABBATE	Internet: su evolución y sus desafíos	143
	BIOMEDICINA	157
GINÉS MORATA	El siglo del gen. Biología molecular y genética	161
JESÚS AVILA Y JOSÉ M. MATO	La biomedicina en el cambio de siglo	171
ALEXANDER KIND Y ANGELIKA SCHNIEKE	Clonación de mamíferos: algo más que una simple oveja	185
JOAN MASSAGUÉ	Hacia una comprensión del cáncer	203
	ECOLOGÍA Y BIOLOGÍA DE LA CONSERVACIÓN	221
CARLOS M. DUARTE	El jardín del Edén amenazado: ecología y biología de la conservación	225
JOHN B. HEYWOOD	Movilidad geográfica en un mundo condicionado por el clima	239
CAYETANO LÓPEZ	Retos actuales de la energía	257
	CAMBIO CLIMÁTICO	273
SERGIO ALONSO	Cambio de clima en el planeta Tierra	277
	ECONOMÍA, FINANZAS Y MANAGEMENT	297
VICENTE SALAS FUMÁS	La economía de la empresa	301
JOAN ESTEBAN	Fronteras de la ciencia económica	315
NATHAN ROSEMBERG	Innovaciones radicales: la visión de un economista	329
	COOPERACIÓN AL DESARROLLO	343
ABHIJIT V. BANERJEE	Por qué es tan difícil combatir la pobreza	347
	ARTES	363
FRANCISCO CALVO SERRALLER	La averiguación artística del futuro	367
LUIS FERNÁNDEZ-GALIANO	La arquitectura del nuevo siglo. Una vuelta al mundo en diez etapas	375
LUIS DE PABLO	¿Fronteras y conocimiento en música? Unos apuntes	391
	BIOGRAFÍAS	404

ciencia, innovación y sociedad: desplazando la frontera de lo posible

FRANCISCO GONZÁLEZ

PRESIDENTE DE BBVA

La obra que el lector tiene entre sus manos, editada por BBVA, está ligada a los Premios Fundación BBVA *Fronteras del conocimiento*, cuya primera edición está a punto de fallarse cuando escribimos estas líneas. Los premios, que tienen una dotación económica de las más elevadas a nivel internacional, se otorgarán anualmente a equipos de investigadores y creadores en ocho categorías, precisamente los ocho campos en los que se articula este libro.

BBVA no obtiene ninguna rentabilidad inmediata de estas iniciativas de apoyo a la generación y difusión de conocimiento. Nuestro grupo no está en el sector farmacéutico, tampoco en el de las tecnologías de la información o las telecomunicaciones, ni en ninguna industria que pueda aprovechar comercialmente, de forma directa, los resultados de la investigación. Sin embargo, se trata de un compromiso que responde a dos vectores centrales de la cultura, la estrategia y la actividad de BBVA: trabajar por un futuro mejor para las personas, y hacerlo promoviendo la innovación, respaldada ésta, a su vez, por el mejor conocimiento disponible. Estamos convencidos de que, de esta forma, BBVA asume una de las funciones que la empresa en general y la gran empresa multinacional, en particular, debe desempeñar en la sociedad global del siglo XXI.

El compromiso de BBVA con el conocimiento: innovación y responsabilidad corporativa

Actualmente existen en el mundo más de 70.000 empresas multinacionales, que representan el 25% de la producción

económica mundial. En las últimas dos décadas, el importe de la inversión extranjera de estas empresas superó al de la ayuda oficial al desarrollo. Puede decirse que estas empresas se han convertido en los instrumentos principales para la construcción de una economía y una sociedad globales, facilitando la difusión de tecnología, de valores y de prácticas comerciales, de gestión más moderna y eficaz en todo el mundo. Además, las grandes empresas tienen una enorme penetración e impacto social, a través de sus empleados, sus clientes, sus proveedores. Por eso, están excelentemente preparadas para ser poderosos catalizadores de la innovación y de la transformación hacia un mundo sostenible.

Las empresas, más que ser parte de los problemas globales, tienen que ser componente esencial de su solución. En el siglo XXI, la empresa responsable, que responde a las demandas legítimas de su entorno, debe asumir compromisos para mejorar las sociedades en las que está presente. Tiene dos motivos muy importantes para hacerlo: la convicción y el interés. La convicción, porque la ética y los valores centrales de su cultura corporativa deben guiar su actuación. El interés, porque, en una sociedad cada vez más informada y exigente, las empresas necesitan una mayor legitimidad para desarrollar con éxito su actividad en el medio y largo plazo. Y porque una sociedad más próspera y estable es, al tiempo, resultado y condición del desarrollo de las empresas.

Si esto es así para todas las empresas, aún lo es más para los bancos. Porque la industria financiera está en el centro de la economía y la sociedad. Su función es ayudar

a las empresas y a los ciudadanos a realizar sus proyectos, prestando servicios fundamentales de pagos, de ahorro y de inversión, además de ofrecer una gama creciente de otros servicios cada vez más especializados. Por eso, la banca es un motor fundamental del desarrollo. Y, por eso, la banca tiene que ser un referente básico de la confianza de todos los agentes sociales, en un doble sentido: en el de atender a los intereses legítimos de sus *stakeholders* (dimensión que recoge la palabra inglesa «*trust*») y, también, en el sentido de la prudencia y competencia profesional (al que se refiere la expresión «*confidence*»). Ética y competencia son dos atributos esenciales, que cada entidad, y el sistema financiero en su conjunto, debe salvaguardar celosamente. Los acontecimientos de los últimos meses muestran los graves efectos para la economía y la sociedad globales —y para las propias entidades financieras— de la quiebra de la confianza en el sector financiero.

Como resultado de su posición central en la economía y la sociedad, la banca se ve afectada de forma plena por los cambios tecnológicos y sociales. Cambian los clientes, sus necesidades, sus expectativas y demandas, los canales y procedimientos por los que prefieren ser atendidos. Y responder a estas exigencias requiere una profunda transformación tecnológica, organizativa y cultural. Una transformación paralela a la que está experimentando la sociedad global para aprovechar el inmenso potencial del avance científico y tecnológico para sostener el crecimiento económico, mejorar el bienestar de la sociedad y restaurar los equilibrios medioambientales rotos en un pasado reciente.

Las materias primas clave de la industria financiera son dos: el dinero y la información. Y el dinero, a comienzos del siglo *xxi*, está en gran medida desmaterializado y se ha convertido en apuntes contables; en definitiva, en información susceptible de ser transmitida instantáneamente y a un coste casi nulo. Por ello, el avance tecnológico da a la banca oportunidades excepcionales para mejorar todos sus procesos y ofrecer más y mejores productos y servicios a un número mucho mayor de personas, con la máxima conveniencia para ellos y a precios mucho más bajos.

Ciertamente, la industria financiera ha cambiado en las últimas décadas para adaptarse, en algunos de sus elementos y operativa, al nuevo entorno tecnológico, pero la profundidad y la amplitud de estos cambios distan mucho de lo que se ha visto en otros sectores industriales y de servicios. Con todo, el peso de los hechos, de las realidades sociales y económicas globales hacen que esa transformación profunda de la industria financiera sea inevitable y cada vez más apremiante.

BBVA quiere ocupar una posición de liderazgo en esa transformación de la industria financiera, a través de una estrategia que se basa en tres pilares: los principios, la innovación y las personas.

Esa estrategia, y la cultura corporativa de la que surge y a la que refuerza, se resumen en nuestra visión: «*BBVA, trabajamos por un futuro mejor para las personas*». Es

decir, trabajamos para las múltiples dimensiones de las personas que son nuestros clientes, nuestros empleados, nuestros accionistas y, también, todos los ciudadanos de las sociedades en las que estamos presentes.

Creemos que nuestro trabajo ayuda a un futuro mejor si trabajamos de acuerdo con firmes principios éticos, de veracidad, honestidad y transparencia. Si situamos a las personas en el centro de nuestra actividad y colaboramos en hacer posibles las aspiraciones de las sociedades en las que estamos presentes. Y, finalmente, si impulsamos la innovación, como vía fundamental para ofrecer más y mejores soluciones, más flexibles e individualizadas y a mejor precio a nuestros clientes, ampliando las posibilidades de acceso a los servicios financieros por parte de más personas. Algo que, al tiempo, generará más valor para nuestros accionistas y permitirá a nuestros colaboradores desarrollar una vida más gratificante.

En ese marco de innovación y responsabilidad corporativa se integra nuestro compromiso con el impulso y la difusión de la educación y la ciencia.

La innovación es un pilar clave de nuestra estrategia y nuestra cultura. En BBVA realizamos un gran esfuerzo de innovación y transformación, con el objetivo de ofrecer mejores soluciones para las personas. Y ello se traduce en una diferenciación respecto al conjunto de empresas de nuestro sector, y en una capacidad superior de creación de valor recurrente, en el medio y largo plazo. No se nos escapa que la ciencia, la investigación y los entornos creativos que las acompañan y son su consecuencia, resultan claves para la innovación de procesos y de productos, de nuevas y más eficaces soluciones a las demandas y los retos de las sociedades del presente.

La responsabilidad respecto a las sociedades en las que trabajamos —actualmente, más de 30 países en varios continentes, y en proceso de ampliación— es parte integral de la estrategia y la cultura de BBVA.

Creemos que nuestra primera responsabilidad es hacer bien nuestro trabajo diario. Mejorar cada día los servicios —la calidad, confiabilidad y precio—, que prestamos a nuestros clientes es la mejor forma de contribuir al desarrollo económico y la estabilidad social.

Pero nuestro compromiso va más allá. Porque queremos contribuir a que ese crecimiento sea sostenible en el tiempo. Por eso, desarrollamos políticas pioneras en el sector en el terreno medioambiental. Y por eso, hemos desarrollado también una estrategia y una política en el terreno social, que incorpora tres líneas básicas de actuación, entre las que el impulso a la educación y el conocimiento tiene un papel central.

La primera de ellas se instrumenta a través de nuestro programa de «Inclusión Financiera», que busca favorecer el acceso a los servicios financieros básicos de pago, ahorro o crédito —servicios que resultan fundamentales para evitar la exclusión económica y social y promover el desarrollo personal— a personas de renta baja para las que la

industria financiera convencional no ofrece un modelo de negocio que sea rentable. En algunas áreas de implantación de BBVA, como Latinoamérica, ése es, precisamente, el segmento mayoritario de los ciudadanos.

Buena parte de esas iniciativas las realiza el propio Banco, que está desarrollando modelos innovadores que, a través del uso intensivo de la tecnología, permiten abaratar drásticamente la producción y distribución de servicios financieros básicos. Además, nuestro Grupo está fuertemente comprometido con el desarrollo de las microfinanzas, para lo que ha creado la Fundación BBVA para las Microfinanzas, entidad sin ánimo de lucro y dotada con 200 millones de euros.

El «Apoyo a la Educación», especialmente en los segmentos de la población de menos recursos, es otra de las líneas de trabajo de nuestro Grupo. Concretamente, el Plan de Acción Social de BBVA en América Latina destina actualmente el 1% de los beneficios de cada banco del grupo en esa región y más del 80% de esos recursos se dedican a diferentes programas educativos.

La tercera línea es, como comentábamos al inicio de estas páginas, la del «Impulso del Conocimiento», que fundamentalmente lleva a cabo la Fundación BBVA a través de diferentes programas de apoyo a la investigación científica y a su difusión, con especial atención a las ciencias sociales, la biomedicina, las ciencias del medio ambiente y las ciencias básicas, además de la cultura (especialmente la literatura española y latinoamericana y la música contemporánea).

Ésta es la línea en la que se inscriben tanto este libro *Fronteras del conocimiento* como los Premios del mismo nombre. Con ellos, queremos contribuir a corregir un déficit general, especialmente marcado en la sociedad española y latinoamericana, de visibilidad y reconocimiento explícito de los múltiples logros de la comunidad científica. Y, con ello, proyectar a la sociedad los avances de la ciencia, la tecnología y algunas de las áreas del arte de nuestra época, el significado de su contribución a la mejora de las posibilidades colectivas y, también, el perfil de quienes con su trabajo de investigación y creación hacen posible la ampliación del espacio de lo conocido y enriquecen el dominio de la cultura.

Constituye una paradoja que la alta estimación social latente de la figura abstracta de los científicos e investigadores, evidenciada en todo el mundo por múltiples encuestas desde finales de la década de 1950, apenas se refleje en ocasiones de celebración y reconocimiento público hacia las personas concretas que logran impulsar de manera decisiva el conocimiento. Algo que contrasta con lo que ocurre con otras profesiones menos centrales. Son pocos los premios internacionales —y ninguno como los galardones del Nobel— que logran, de tiempo en tiempo, hacer visibles para la sociedad la identidad de un selecto grupo de investigadores y creadores. Los Premios Fronteras del conocimiento quieren contribuir también a ese objetivo de acercamiento y valoración social de la cultura científica del presente.

Los nuevos premios que BBVA impulsa a través de su Fundación presentan un perfil diferente al de otros galardones importantes. Pretenden reconocer e incentivar la investigación y la creación cultural, en especial aquellas contribuciones de amplio impacto por su originalidad, implicaciones teóricas y conceptuales y su traslación a innovaciones particularmente significativas. La denominación de estos premios quiere significar tanto el trabajo de investigación capaz de ampliar el ámbito del conocimiento —desplazando hacia delante, de manera continua, la frontera de lo conocido—, cuanto el encuentro y solapamiento entre áreas disciplinares. Específicamente las ocho áreas objeto de reconocimiento y representadas en los capítulos de este libro, son las siguientes: Ciencias básicas (Física, Química, Matemáticas), Biomedicina, Ecología y Biología de la conservación; Cambio climático; Tecnologías de la Información y la Comunicación; Economía, Finanzas y Gestión de Empresas, Cooperación al desarrollo y Arte (Música, Pintura, Escultura, Arquitectura).

Además del número de áreas y el contenido de las mismas, hay otros elementos que dan a estos premios un perfil específico. En primer lugar, el reconocimiento del carácter interdisciplinar del conocimiento en las últimas décadas del pasado siglo y en el presente. En segundo lugar, atienden al hecho de que muchas de las contribuciones decisivas al conocimiento de nuestra época son resultado de la colaboración de grandes equipos de investigadores, lo que lleva, a diferencia de otros premios, a aceptar la posibilidad de hacer objeto del galardón a equipos, sin restricción del número de sus integrantes, o a un número abierto de investigadores que de manera paralela hayan convergido en hacer posible un determinado avance. En tercer lugar, se establece una categoría que premia la creación de excelencia en cuatro ámbitos de la cultura particularmente innovadores y que influyen de manera significativa en la sensibilidad de una época, como son la música contemporánea, la pintura, la escultura y la arquitectura. En cuarto lugar, se reconocen esfuerzos de tipo diverso (desde la investigación hasta las actuaciones e iniciativas prácticas) en dos cuestiones centrales de la sociedad global del siglo XXI: el cambio climático y la cooperación al desarrollo de amplias zonas del planeta, o lo que es lo mismo, la eliminación de la pobreza y la marginación.

Estos premios responden a la visión de BBVA sobre el conocimiento y la innovación en las condiciones de la sociedad global del presente, visión que esbozamos en las restantes páginas de este capítulo, como una presentación general del sentido de este libro.

Sociedad del conocimiento y sociedad global

El término «sociedad del conocimiento» y otros emparentados con él, como «sociedad de la información» y «economía del conocimiento», surge en los años 60 del pasado siglo. Todos ellos tienden a designar el mismo fenómeno —o distintas facetas del mismo fenómeno—: la emergencia de una

sociedad (una economía) en la que los avances científico-tecnológicos y, en particular, los de las tecnologías de la información y las telecomunicaciones, y su rápida difusión a la sociedad, se convierten en elementos centrales de la actividad económica, transformando en paralelo y de manera profunda la cultura y los modos de vida de las personas.

Posiblemente, el primer autor en utilizar el concepto de «sociedad de la información» fuera el economista Fritz Machlup, quien en 1962 constató cómo el número de empleados dedicados a la manipulación de información era mayor que el de los que desarrollaban algún tipo de actividad física o manual.

Por su parte, la expresión «sociedad del conocimiento» fue utilizada por primera vez por Peter Drucker en su libro de 1968 *The Age of Discontinuity*, postulando que el recurso básico para la producción de riqueza en nuestro tiempo reside en el conocimiento y en la productividad de éste. Posteriormente, este mismo influyente autor subrayaría las profundas transformaciones sociales que esto implicaba y la estrecha vinculación entre «conocimiento» y «globalización»: el auge de las tecnologías de la información y la telecomunicación permitían —en opinión de Drucker— que empresas, productos y consumidores pudieran trascender las fronteras nacionales e impulsar la emergencia de un gran mercado global (Drucker 1994).

En las décadas transcurridas desde las contribuciones iniciales de Machlup y Drucker, las tendencias detectadas se han reforzado hasta un extremo difícil de anticipar. No es extraño por ello que en esas décadas se propusieran una serie de modelos sociológicos y económicos acerca del tránsito desde la sociedad industrial a la sociedad posindustrial, incorporando los atributos principales de la sociedad de la información. Dos de los autores más conocidos de esos modelos son A. Touraine y, especialmente, Daniel Bell.¹ Ha habido otras muchas caracterizaciones similares de las sociedades del último tercio del siglo pasado. Las etiquetas más conocidas, compiladas por Beniger (1986), destacan la importancia de la base *tecnológica* —particularmente las «tecnologías de la información»— para modelar la estructura de las sociedades avanzadas: *sociedad computerizada, era de la información, communications, sociedad posindustrial, revolución electrónica, sociedad telemática, sociedad cableada, era computacional, micro-milenio, tercera ola, era de la información*. A esa lista podrían añadirse otras expresiones emparentadas, como las de *revolución del control*, debida al propio Beniger, *High Tech Society* o, dando un salto hasta la última década del siglo xx, «*Network Nation*», «*Virtual Community*», «*the Network Society*», «*Cybersociety 2.0*».

Avance tecnológico, cambio cultural e innovación

Las interacciones entre las tecnologías electrónicas, los nuevos materiales, la informática y las telecomunicaciones, además de los desarrollos en curso en los campos de la nanotecnología y de la biotecnología, han hecho po-

sible reemplazar la base tecnológica que sostuvo varias décadas de crecimiento ininterrumpido desde finales de la Segunda Guerra Mundial hasta casi el cierre del siglo anterior. Uno de los componentes esenciales del actual impulso tecnocientífico —la asociación de informática y telecomunicaciones— presenta la particularidad de afectar a procesos y productos de todos los sectores económicos sin excepción, abarcando además todo el ciclo económico, desde el diseño a la comercialización, pasando por la ingeniería y la producción. Además de esos impactos económicos directos de las tecnologías de la información, sus efectos fundamentales pueden medirse también, desde hace al menos dos décadas, en una larga serie de áreas como la propia práctica del trabajo científico, la enseñanza, la sanidad, el ocio, las prácticas asociativas (emergencia de grupos de interés, asociacionismo electrónico y comunidades «virtuales») y el ámbito de la cultura.

Los progresos en la tecnología de la información en el último medio siglo han sido formidables. La llamada Ley de Moore (en realidad, una observación empírica, que establece que la capacidad de almacenamiento y proceso de información se dobla cada 18 meses), se ha mantenido desde que Gordon Moore la formulara en los años 60. Pero no se trata sólo de esto, sino, también, y muy especialmente, de la difusión universal de los ordenadores personales y el desarrollo de Internet, una plataforma cuyo enorme poder reside en la combinación de la *información* que alberga y recrea de continuo, y su condición de *red*. Precisamente, la llamada Ley de Metcalfe, establece que el valor de una red aumenta en relación al cuadrado del número de personas conectadas a ella.

Es claro que uno de los factores explicativos de la rápida difusión de la informática reside en los avances científico-tecnológicos, que han hecho posible la mejora sostenida en prestaciones, en paralelo a la baja de su coste. Pero la trayectoria del computador desde mediados de los años 40 del pasado siglo hasta el presente ha estado afectada también por dimensiones sociales, algunas tan intangibles como el *modo de ver* al computador: cuáles son las posibilidades que abre y cómo se concibe la «cohabitación» de procesadores humanos y procesadores electrónicos, para decirlo con la plástica imagen del Premio Nobel Herbert Simon (1985). Merece la pena referirse, siquiera sea brevemente, a esos dos aspectos, tecnológico y social, por cuanto ilustran la complejidad de los procesos de innovación.

Hoy tomamos como algo obvio que el computador es una tecnología de propósito «universal» («The Universal Machine») y, por ello, situada en plano muy distinto del ocupado por las máquinas emblemáticas de la revolución industrial, «dedicadas» por diseño a una o, como mucho, a unas pocas tareas, prefijadas y especificadas de antemano. Vemos y usamos el computador como una tecnología susceptible de convertirse en soporte y amplificador de una larga serie de funciones mentales y tareas especializadas,

¹ Daniel Bell señaló en el Prólogo de 1976 a la segunda edición de *The Coming of the Postindustrial Society* su incomodidad con las etiquetas «sociedad de la información» y «sociedad del conocimiento» o «sociedad de servicios», que sólo remitirían a aspectos parciales del modelo de sociedad emergente. Pero el foco del análisis y el tenor literal del argumento de Bell recae tan claramente en la dimensión technoindustrial de la nueva sociedad, que autoriza su inclusión, dentro del grupo de autores del modelo de la sociedad de la información.

en ampliación permanente. Esta versatilidad permite trascender la función a la que la propia etiqueta lingüística alude (*computing* o «cálculo»), para abrir paso a funciones tan variadas como el tratamiento integral de información cuantitativa o categorial (cualitativa), la creación de imágenes e incluso de «mundos virtuales» y muchas más, entre las que destaca el ser, en interacción con las telecomunicaciones, una plataforma de comunicación flexible y robusta, soportando todos los formatos, desde la voz al texto, las imágenes y el vídeo, etc., abarcando además al conjunto del planeta y dando forma así a «Global Networks» (Harasim 1993).

Tan viva es la percepción de la brecha con el pasado reciente abierta por la difusión universal del computador y el espacio web tejido con él, que hay que recurrir a los historiadores de la tecnología para darse cuenta de que el modo dominante de concebir los primeros computadores electrónicos, incluso entre la mayoría de sus arquitectos, era el de una tecnología altamente especializada, llamada a ocupar un nicho limitado en unas pocas grandes organizaciones: corporaciones, ejército e instituciones científicas. Paul Ceruzzi, colaborador en esta obra, ha destacado que incluso a la altura de 1951 se estimó que todas las necesidades computacionales de Estados Unidos quedarían satisfechas con cuatro o cinco computadores (Ceruzzi 1986). Una visión tan restrictiva acerca de la gama posible de aplicaciones y destino del computador se explica por factores tecnológicos y culturales.

Comenzando por el plano tecnológico, el análisis de los procesos y la historia de las innovaciones científico-tecnológicas han mostrado que para maximizar las potencialidades de una determinada tecnología se requiere la confluencia de distintos desarrollos tecnológicos (*clusters* de distintos avances)(Freeman y Soete 1997). Para Rosenberg (1976), la longitud del intervalo entre la fecha de una «invención» y su difusión como «innovación» es, en gran medida, función del tiempo transcurrido para llevar a término actividades inventivas *adicionales*, orientadas al refinamiento del diseño original, teniendo presente los requerimientos del usuario final.

En el caso que nos ocupa, sin la revolución microelectrónica no cabría explicar el fenómeno de la producción seriada y masiva de computadores y su difusión abarcando desde la gran empresa al hogar, tampoco la aparición de lo que el desaparecido científico del MIT, Michael Dertouzos, llamó *hidden computers*, computadores «ocultos», insertos en una amplia serie de productos y tecnologías (máquinas, herramientas avanzadas, automóviles, electrodomésticos, equipos de imagen y sonido, relojes y un largo etcétera), mejorando sus prestaciones, incrementando su fiabilidad y permitiendo un considerable ahorro de materiales e, incluso, el diagnóstico y reparación remotos (Dertouzos 1984).

En la dimensión cultural, lo ocurrido con el computador en su periodo de despegue se asemeja al caso de otros avances tecnológicos radicales en la fase de su emergen-

cia: la proyección sobre la nueva tecnología de las imágenes de la herramienta o tecnología a la que «reemplazaba». Así, el coche fue visto como «un carruaje sin caballos», en lugar de como el inicio de una «era de auto-movilidad», la radio como un «telégrafo sin hilos», e, incluso Graham Bell, durante un corto espacio de tiempo, consideró el teléfono como un medio de difusión de mensajes desde una central (al modo de la radio), en lugar de como una herramienta conversacional. No tiene, por tanto, nada de extraño que el computador electrónico fuera concebido como una potente herramienta llamada a reemplazar con ventaja a los únicos «computadores» existentes en la época: las personas dedicadas, armados con lápiz y papel o con el concurso de tabuladoras mecánicas, al cálculo de tablas utilizables para distintas aplicaciones (navegación, seguros).

Ciertamente, la posibilidad misma del computador *personal* ha requerido de multitud de avances en los lenguajes de programación, interfases, sistemas operativos, aplicaciones, que han eliminado prácticamente las barreras de entrada para su uso. Pero, también, factores de naturaleza cultural, como «visiones» acerca de los usos del computador y sus modos de relación con el usuario final, han resultado fundamentales para su pauta de difusión masiva. Visionarios como Vannevar Bush, Douglas Engelbart, el proyecto Augmented Human Intellect del Stanford Research Institute (SRI), la comunidad de investigadores en Inteligencia Artificial, el legendario centro de investigación de Xerox en Palo Alto, el programa de investigación en «trabajo cooperativo asistido por ordenador» (CSCW), así como la implementación de algunos de los principios guía de esa(s) comunidad(es) en el arranque de la empresa Apple, hicieron cristalizar una visión del computador como tecnología para «aumentar» (en lugar de reemplazar) las capacidades de los seres humanos, y para ensanchar las posibilidades de comunicación y colaboración en el grupo-de-trabajo y, más allá de éste, entre grupos sociales e individuos.

El salto desde la decena de grandes computadores de la década de 1940, aplicados a la ejecución de unas pocas tareas especializadas por un reducido segmento de científicos e ingenieros, a los millones de microcomputadores de propósito universal, accesibles incluso para los niños, en la década de 1990, representa un hito fundamental en la historia de la tecnología y de sus impactos sociales. Pero en el periodo más cercano, el acercamiento y fusión de las telecomunicaciones y de la informática, cristalizada en el crecimiento exponencial de las «redes» y de Internet, ha abierto una nueva y más decisiva línea divisoria en el desarrollo de la tecnología, así como un espacio sin precedentes para la experimentación social.

Todo ese complejo proceso ilustra cómo los procesos de innovación radical, para resultar exitosos, requieren de la interacción de numerosos avances, desde los estrictamente tecnológicos a los de naturaleza social y cultural. Y cada uno de ellos tiene su propio tiempo de desarrollo, lo que dificulta

la anticipación precisa del «cuándo» e incluso del «qué», la emergencia de una innovación con efectos significativos.

¿Hacia una verdadera sociedad del conocimiento?

Entre los efectos sociales del proceso de innovación de la informática y las telecomunicaciones, quizás el principal de todos sea el que los ciudadanos vivimos hoy en una sociedad de la información. Tenemos acceso a un universo de información en perpetua expansión. Contamos con una tecnología cada vez más potente, ampliamente accesible, para crear y acceder a esa información, transmitirla, tratarla, analizarla, relacionarla, y, eventualmente, convertirla en conocimiento y aplicarla a la resolución de problemas. Con ello, en poco más de tres décadas se ha invertido una pauta secular caracterizada, en primer lugar, por la penuria absoluta de información y, en segundo pero no menos importante lugar, por la concentración de la misma en una minoría de la sociedad.

Desde luego, nuestra sociedad es, con mucho, la que más conocimiento genera y ha acumulado en la historia de la Humanidad. En 1990, David Linowes afirmaba que hasta la mitad del siglo XVIII no se dobló el conocimiento que existía en la época de Cristo. Volvió a doblarse en los siguientes 150 años —hasta el inicio del siglo XX— y de nuevo se dobló en 50 años —hasta la mitad de ese siglo—. En nuestros días, el volumen de conocimiento se dobla cada cuatro o cinco años.

Sin embargo, una verdadera sociedad del conocimiento no es todavía más que una aspiración, una meta hacia la que avanzamos, pero que resulta elusiva. Y esto es así por dos razones fundamentales: primera, porque gran parte de la Humanidad continúa fundamentalmente ajena a todos estos desarrollos. Como señala Janet Abbate en su artículo en este libro, más de la mitad de la población de los países desarrollados utilizaba Internet en 2005, frente al 1% en los 50 países menos desarrollados. Hoy existe una inmensa brecha digital (*digital divide*), que frena las posibilidades colectivas abiertas por el computador y la web y que demanda atención urgente por parte de agentes públicos y privados, si se quiere acortar el diferencial entre sociedades avanzadas y países que no han logrado entrar en la senda del crecimiento sostenible.

La segunda razón por la que la sociedad del conocimiento es todavía más objetivo que realidad empíricamente observable, es que la inmensa mayor parte de la información disponible es una enorme masa indiferenciada de datos cuantitativos y de información categorial o cualitativa. La estructura de gran parte de esa información sigue oculta (no conocemos las relaciones internas entre las múltiples piezas que la integran), no disponemos de explicaciones articuladas que les den coherencia global y, en definitiva, no dominamos una «tecnología» fundamental: cómo convertir esos datos en conocimiento y cómo convertir una parte significativa del conocimiento en innovación, en nuevas aplicaciones útiles para la vida de las

personas y para resolver los grandes problemas de orden planetario. Un ejemplo claro de esta situación es la explosión reciente de la información genética (desciframiento del genoma humano), y las dificultades para su interpretación y traslación a terapias nuevas y más eficaces. Pero hoy se trabaja de manera intensa en la construcción de algoritmos estadísticos y metodologías capaces de ayudar a descifrar el sentido de volúmenes gigantescos de información de naturaleza diversa. De su éxito dependerá, en gran medida, nuestra capacidad para transformar la información en conocimiento y su traslado, vía innovación, a satisfacer necesidades y demandas en multitud de planos (Hastie, Tibshirani y Friedman 2003).

Sin duda, la revolución tecnológica y el rápido aumento del conocimiento han dado lugar a una fase fuertemente expansiva del crecimiento económico mundial. Sin embargo, la distribución de esos beneficios se ha revelado muy desigual. La capacidad productiva actual del planeta permite sustentar a una población triple que la de mediados del siglo XX. Han mejorado también los niveles de vida en gran parte del mundo y la pobreza extrema se está reduciendo, no sólo en términos relativos al total de la población, sino también en términos absolutos (Naciones Unidas 2007).

Sin embargo, han aumentado las desigualdades en, prácticamente, todos los países del mundo, y entre las áreas más avanzadas y las más desfavorecidas. No es extraño por ello que proliferen las reflexiones sobre el malestar y los problemas de la globalización y que, en ocasiones, predomine la visión de los problemas más que la de las oportunidades de la globalización y el cambio científico-tecnológico.²

Además, el acelerado crecimiento de la población y de la actividad productiva esta acarreado muy serios problemas de sostenibilidad medioambiental, asociados a la sobreexplotación de los recursos naturales terrestres y marítimos, la escasez de agua dulce, la pérdida acelerada de biodiversidad (especies y hábitats) y el cambio climático, con los impactos consiguientes que ello pueda tener sobre el asentamiento de la población y la economía en las próximas décadas (temas fundamentales sobre los cuales presentamos también en este libro contribuciones de destacados especialistas).

Con todo, hay razones para el optimismo. Las ciencias básicas siguen conquistando capas cada vez más profundas del conocimiento de las estructuras y los procesos físicos, químicos y biológicos. Y de ello se siguen consecuencias de todo orden, desde las estrictamente cognitivas (mejor conocimiento) a las tecnológicas (nuevos instrumentos de satisfacción de necesidades). Nos encontramos en el inicio de la evolución de tecnologías potentísimas y muy versátiles, entre ellas Internet, cuya tercera generación abrirá muchas más posibilidades de cooperación e incorporación activa de todos al espacio electrónico. Estamos meramente escurriendo la superficie de

2

En 1998, la socióloga de la Universidad de Chicago Saskia Sassen publicaba *Globalization and its Discontents*. Nueva York: The New Press, 1998, y cuatro años más tarde, bajo idéntico título, lo hacía el Nobel de Economía, Joseph E. Stiglitz. *Globalization and its Discontents*. Nueva York-Londres: W. W. Norton & Company, 2002.

una gigantesca veta de riqueza y bienestar para la Humanidad. Y ésta es, seguramente, la revolución tecnológica más «democrática» que haya existido nunca. No sólo por el dato contextual obvio de ocurrir cuando la democracia es el régimen político imperante en la mayor parte del mundo —lo que, sin duda, potencia sus efectos positivos— sino, también, porque es la revolución que, a pesar de las barreras y las limitaciones apuntadas, más rápidamente se está difundiendo por todo el planeta, siendo la que plantea menos obstáculos para la incorporación a ella de las personas de todas las partes del mundo.

¿Cómo acelerar la generación de conocimiento?

Este marco, en que coexisten problemas serios y cada vez más apremiantes, junto con enormes potencialidades de avance científico y tecnológico, sitúa la cuestión de los mecanismos y procedimientos más apropiados para impulsar la generación y difusión del conocimiento como una de las cuestiones clave para el futuro de la Humanidad.

Tradicionalmente, los catalizadores principales para impulsar la generación del conocimiento han sido dos: el provecho económico, motor de los agentes privados; y la guerra, motor de iniciativas y programas de los gobiernos.

Las necesidades bélicas como impulsoras de avances científicos y tecnológicos, alcanzaron una eficacia sin precedentes en el contexto de la Segunda Guerra Mundial. La ciencia y las tecnologías subatómicas, las telecomunicaciones, el radar, la microelectrónica y el ordenador recibieron ese impulso del esfuerzo bélico. Pero otras áreas conocieron también avances fundamentales, en particular en la medicina, la farmacología, la psicología y la investigación operativa.³

El *push* que la Segunda Guerra Mundial dio a la investigación científica y tecnológica se vio reforzado en las últimas décadas del siglo pasado con una participación creciente del sector privado, respondiendo a las mayores oportunidades de explotación comercial de los avances en la investigación.

En los últimos cincuenta años ha habido cambios profundos en la forma en la que se genera y se utiliza el conocimiento. En paralelo a un crecimiento sin parangón del conocimiento científico, los gobiernos, las empresas, las instituciones científicas han venido debatiendo las formas más eficaces de traslación de los avances científicos a la competitividad de empresas y los países, en definitiva, a la mejora de las posibilidades colectivas.

Durante décadas, el enfoque dominante fue el apoyo sin reservas por parte de los poderes públicos y, en menor medida, de algunas grandes corporaciones a la llamada investigación básica, bajo el supuesto de que de esos conocimientos, antes o después, se derivarían aplicaciones prácticas muy difíciles de anticipar, particularmente en el caso de aquellas innovaciones de tipo más radical, las más difíciles de imitar para países y empresas competidoras. Ese modelo aparecía más o menos emparentado con la

doctrina de Vannevar Bush y la experiencia de los años de la Segunda Guerra Mundial.⁴

Pero al menos desde los años 80 del pasado siglo la insatisfacción conceptual y práctica con ese enfoque heredado se hicieron patentes. Así, por ejemplo, un influyente informe interdisciplinar, publicado bajo el título *Made in America* (Dertouzos, Lester y Solow 1989), preparado por la Comisión de Productividad Industrial del MIT, entre cuyos miembros se encontraba el Nobel de Economía Robert Solow, quiso explicar la paradoja de que Estados Unidos contara, por una parte, con la ciencia básica más avanzada y los científicos y tecnólogos mejor formados y, por otra, hubiera disminuido su capacidad de traducir ese diferencial cognoscitivo en innovación, en comparación con lo ocurrido en las décadas subsiguientes al término de la Segunda Guerra Mundial. En tanto que Japón, concentrándose en áreas de investigación tecnológica de carácter más aplicado, parecía mostrarse capaz de capturar mercados de productos inequívocamente americanos (como la electrónica de consumo), tenía un tiempo medio de llegar al mercado sensiblemente más corto en sectores emblemáticos de la industria americana (el del automóvil) y alcanzaba niveles de calidad muy superiores.

Esas dificultades distaban de ser privativas de Estados Unidos, como parecieron creer los autores del mencionado informe. Con casi una década de retraso, la Comisión Europea (re)descubrió el mismo problema y no dudó en etiquetarlo como «paradoja europea» en *El libro verde de la innovación* (1995) y en el proyecto *Made in Europe* emparentado con él.

De ese tipo de análisis ha resultado —no sin críticas por parte de sectores de la comunidad científica— un cambio de perspectiva en la elaboración de la política científica, consistente en la reducción del peso de la investigación orientada al avance del conocimiento (lo que se conoce como «investigación pura» o básica), sin búsqueda directa de aplicaciones prácticas. Ese «tipo» de ciencia ha tenido que ir cediendo o, al menos, teniendo que compartir recursos financieros y humanos con la llamada «investigación estratégica», objeto de planificación externa (por agencias e instituciones públicas) al servicio de la satisfacción de objetivos socioeconómicos (destacando de manera nítida entre ellos, la mejora de la competitividad de las economías nacionales, las políticas de defensa y las de salud).

En el ámbito de las grandes empresas se ha ido imponiendo esta última visión, orientada a un alineamiento estricto entre programas de investigación y objetivos económicos. Esto ha llevado, en no pocos casos, a reducir el peso, o incluso a prescindir por completo del propio laboratorio industrial, sacando éste al mercado. O, en otros casos, a sustituir el departamento de I+D por contratos con centros públicos o institutos privados dedicados exclusivamente a la investigación. La incertidumbre en cuanto a los resultados alcanzables y las dificultades para apropiarse en exclusiva de los resultados derivados

3 Sobre la interacción entre necesidades militares y el desarrollo de la tecnología, pueden verse la obra clásica de William H. McNeil, *The Pursuit of Power. Technology, Armed Forces, and Society since A.D. 1000*. Chicago: University Press, 1982, y la obra editada por Merrit Roe Smith para el caso emblemático de Estados Unidos, *Military Enterprise and Technological Chance. Perspectives on the American Experience*. Londres: MIT Press, 1987.

4 Para un análisis crítico de las fortalezas y limitaciones del modelo de apoyo a la ciencia vinculado al ingeniero del MIT, Vannevar Bush, puede verse Claude E. Barfield, ed. *Science for the 21st Century. The Bush Report Revisited*. Washington: the American Enterprises Institute Press, 1997.

de la investigación, particularmente de la básica, han pesado en esta reconsideración del tamaño, papel y formas del gasto en I+D.

Por su parte, las administraciones públicas de las sociedades avanzadas dieron por agotado, desde los años 80 del pasado siglo, el modelo posbélico de desarrollo de la ciencia básica.

Pero ocurre que los supuestos del modelo de dirección estratégica de la ciencia distan de estar sólidamente fundamentados. La historia de la tecnología y la de la innovación muestran el curso tortuoso que lleva del progreso puramente teórico a la mejora o la introducción de nuevos procesos y productos y viceversa.

Ciertamente, resulta muy difícil predecir y gestionar la traducción de avances teóricos en nuevas aplicaciones. Pero, por otra parte, los vínculos entre teoría y aplicación práctica se han multiplicado y sus raíces se han hecho más profundas, obligando a romper esquemas simplificadores acerca de lo que es útil y de lo que —se supone— «sólo» contribuye a un mejor conocimiento de la realidad. En los países y regiones con un entorno cultural e institucional comprometido con la excelencia y la innovación, los agentes públicos y privados comparten la percepción de que la economía y sociedad descansan cada vez más en una infraestructura de «intangibles», de teorías, información y conocimiento científico, donde la actividad científica y las estrategias empresariales tienen un amplio área de solapamiento, en continua redefinición.

Los países europeos que, por lo general, incorporan con retraso conceptos y experiencias del otro lado del Atlántico, necesitan prestar más atención a lo que, de verdad, ha tenido lugar en Estados Unidos: la literatura y la evidencia empírica muestran que la investigación científica financiada con fondos públicos ha desempeñado el papel principal en la innovación industrial en Estados Unidos.

Lo que se desprende de las pautas de innovación en Estados Unidos, es la necesidad de incrementar en este lado del Atlántico el esfuerzo público y privado en actividades de I+D, al tiempo que se promueve la ciencia y la tecnología de excelencia y, sobre todo, se introduce la cultura de «mercado», de competición abierta y esfuerzo sostenido por la excelencia (universidades y centros de investigación estratificados por su capacidad de contribución al conocimiento y la innovación), de movilidad e interacción entre investigadores y el sector privado, rediseñando interfases de comunicación eficaz entre instituciones dedicadas a la creación y transmisión de conocimiento y el mundo de la empresa. Un programa cuyo desarrollo requiere del concurso decidido y la coordinación de las administraciones públicas, desde las de ámbito europeo a las regionales, pasando por las de carácter nacional.

Hay que renegociar el anterior «contrato implícito» entre universidades, industria y administración, redefiniendo qué es lo que puede esperar y dar cada una de esas instituciones a las otras. Tal y como han señalado dos destacados expertos en innovación, los profesores Rosenberg y

Nelson, hay que modificar el *status quo*, pero las reformas deben estar basadas en una consideración cuidadosa de la especialización funcional de unas y otras instituciones, buscando una mejor división del trabajo entre todos los agentes del sistema de innovación.

Lo que parece claro es, en todo caso, la necesidad de impulsar una tupida red de relaciones entre industria y universidad, que puede tomar formas varias, desde el acceso fluido del personal de I+D de la empresa a sus colegas universitarios (y a la inversa), al desarrollo de instituciones especializadas a medio camino entre empresa y centros de investigación públicos, pasando por la financiación pública de líneas de investigación cuyo propósito sea la mejora de la competitividad, supervisadas por comités asesores de composición mixta universidad-empresa-administración. Con independencia de unas u otras fórmulas, lo realmente importante es crear redes de transmisión dinámica de información y señales, la generación de confianza e intercambio de conocimiento tácito (difícilmente codificable) entre los distintos agentes, derribando las barreras observables en gran parte de Europa entre, por decirlo sintéticamente, universidad y empresa.

Interacciones entre ciencia y tecnología

Los resultados de la investigación científica y la innovación tecnológica se han hecho cada vez más presentes en todos los aspectos de la actividad y de la vida humanas, hasta el punto de que están generando, como señalara Peter Drucker, «mucho más que una transformación social, un cambio en la propia condición humana» (Drucker 1994). Ello se traduce en una interpenetración y una fertilización cruzada crecientes entre la investigación científica, la innovación, las actividades productivas y los modos de vida de las personas. Y en una drástica reducción del tiempo entre el descubrimiento científico y la explotación comercial de sus resultados (Mowery 1989).

La ciencia y la tecnología avanzan, cada vez más, como resultado de la confluencia y la interacción entre disciplinas «clásicas» y de la emergencia de otras nuevas, así como del solapamiento y de la influencia cruzada entre ciencia y tecnología, dejando como algo obsoleto la clásica discusión de si la tecnología depende del previo conocimiento científico, si éste se beneficia de aquélla o, si marchan de manera independiente. En el siglo xx, y muy en particular en su segunda parte, las relaciones ciencia-tecnología y de ambas con la sociedad han cambiado de manera fundamental. Las empresas y los sectores industriales, así como las demandas sociales en campos como la salud, la energía, la agricultura y la alimentación, el transporte, el medio ambiente, la desigualdad y la pobreza son fuentes o señales para la ciencia, reclamando la potencia analítica que sólo la investigación científica puede proporcionar y las soluciones eficaces y eficientes propias del dominio de la tecnología —el ámbito que el Nobel de Economía Herbert Simon etiquetara como «las ciencias de lo artificial» (1996).

Este complejo marco del presente es, también, el que ayuda a explicar el peso creciente de la cooperación multidisciplinar en la investigación científica contemporánea, así como el hecho de que la mayor parte de la investigación científica se desarrolle por equipos muy numerosos, integrados por investigadores afiliados con distintas instituciones y basados en diferentes puntos del mundo. Las innovaciones en las telecomunicaciones e Internet permiten la participación activa y simultánea en un determinado proyecto de investigadores de todas las partes del mundo, incluyendo —y esto es bastante esperanzador— los de las regiones menos ricas y avanzadas del planeta.

Arquitectura institucional y entorno cultural de la ciencia

La ciencia es hoy una actividad marcadamente social y altamente institucionalizada, que si bien sigue requiriendo de la creatividad y el riesgo individual, se desarrolla cooperativamente en marcos organizativos especializados y en un entorno social del que obtiene no sólo medios humanos y materiales adecuados, sino también señales (apreciación de la ciencia, prioridades de la investigación) e influencias conceptuales y culturales, desde las provenientes de campos próximos, a las propias de las Humanidades y el conjunto de la cultura (las *worldviews* activas en una sociedad).

La importancia del alineamiento y la interacción positiva entre todos esos elementos se ha hecho particularmente crítica en las últimas décadas. Y, como explica Nathan Rosenberg en su excelente artículo en este libro, un reto crucial es la adaptación del marco institucional (*institutional setting*) de la ciencia y la investigación a nuestra sociedad global.

En este sentido, hoy se plantean nuevos retos organizativos para el desarrollo de una investigación que es crecientemente interdisciplinar, multipolar —incluso deslocalizada—, fuertemente cooperativa y que mantiene un grado de interacción creciente con su medio social.

¿Cómo desarrollar investigación interdisciplinar en universidades divididas en departamentos definidos por disciplinas específicas, y por investigadores y científicos que —al menos en el mundo académico— asignan una gran importancia al hecho de trabajar en un campo reconocido?, ¿cómo combinar los marcos y estructura disciplinar del conocimiento, bien definidos y asentados en razones y tradiciones teóricas, con institutos y centros interdisciplinares, más cercanos al abordaje de retos prácticos?, ¿cómo conciliar los intereses de los gobiernos y las agencias públicas nacionales, que son una parte integral del entramado científico, con la configuración de equipos multinacionales muy flexibles y cambiantes?, ¿cómo salvaguardar los incentivos de las empresas a destinar recursos a la investigación, en proyectos con múltiples participantes y en un contexto en el que información vital puede ser divulgada instantáneamente *urbi et orbe*? Y, finalmente, ¿cómo asegurar que todo

este entramado institucional se orienta a la resolución de problemas de interés general y contribuye efectivamente al bienestar de las personas, de forma que el avance científico y tecnológico no se traduzca en un aumento de las desigualdades o en mayores problemas de sostenibilidad global?

La mera enumeración de estos desafíos hace ver que, en gran parte, demandan respuestas del campo de las ciencias sociales y del comportamiento y, en particular, de las tecnologías «blandas» de la organización y los incentivos, así como del estudio de la cultura y las actitudes.

Pero además del rediseño de la arquitectura institucional de la ciencia y la tecnología, de las políticas públicas para su promoción y de la gestión estratégica de las actividades de I+D y de innovación por parte de las empresas, se requiere que el apartado de los intangibles, los valores y las percepciones, en definitiva la envolvente cultural de la ciencia sea sensible a ésta, opere como impulso y como vector de orientación.

Reconciliando ciencia, tecnología y sociedad

Una percepción social empática respecto a la ciencia es crucial en, al menos, tres aspectos fundamentales. Primero, para que la ciudadanía impulse, con el peso de su opinión, de su voto, incluso de su poder de compra, las políticas públicas y las decisiones de las empresas privadas de apoyo e inversión en educación y en investigación. Premiando la innovación (*the high road*) y desincentivando la ruta competitiva basada meramente en costes bajos, más que en alto valor añadido. Segundo, para atraer capital humano hacia la ciencia, para que los jóvenes con talento se sientan motivados a emprender una exigente, pero también apasionante carrera investigadora, recompensada económica y simbólicamente. Finalmente, la «apropiación» intelectual y cultural de la ciencia por la sociedad es crucial para la propia creatividad científica y para el aprovechamiento e integración eficaz de los nuevos desarrollos en el tejido social. En definitiva, la envolvente que, a propósito del caso de Estados Unidos, ha sido etiquetada por el historiador de la tecnología, Thomas Hugues, como «entusiasmo tecnológico» es decisiva para el avance del conocimiento y, con el, de la sociedad que lo impulsa y acoge (Hughes 2004).

Podríamos sentirnos tentados a pensar que tras varias décadas en las que la ciencia y la tecnología han hecho una contribución abrumadora al progreso económico y al bienestar de las personas, la consideración general de la ciencia como un factor inequívocamente positivo debería estar firmemente consolidada. Pero tal y como destaca Gerald Holton en su excelente ensayo en este libro, la valoración social de la ciencia ha estado históricamente sujeta a fuertes oscilaciones, y ese estatus social deseable de la ciencia no está garantizado. Tras el progresismo optimista que resurgió tras la Segunda Guerra Mundial, ejemplificado en el célebre informe de Vannevar Bush, *Science, the Endless Frontier*, encargado por el presidente

Roosevelt y publicado en 1945, en la última parte del siglo xx han emergido voces y movimientos críticos con el papel de la ciencia en nuestra sociedad, que recuperan motivos de la resistencia romántica ante la ciencia y la modernización (Marx 1988). Se atribuyen al progreso científico y tecnológico efectos negativos en relación con el desarrollo de armas de inmenso poder destructivo, el deterioro del medio ambiente, la desigualdad dentro de cada sociedad y entre diferentes partes del mundo e, incluso, la configuración de una sociedad y una cultura globales indiferenciadas, deshumanizadas, excesivamente materialistas y privadas de valores morales.

Ciertamente, la preocupación por estas cuestiones y, en particular, por la situación de serio deterioro global del medio ambiente es no sólo legítima, sino compartida por muchísimos ciudadanos. El propio Leo Marx, historiador en el MIT de la cultura americana, ha hecho notar que la creencia en el progreso característica de la cultura moderna euroamericana ha resultado erosionada en las tres últimas décadas del pasado siglo, principalmente por el pesimismo acerca del papel de los seres humanos en la naturaleza y la percepción de que el sistema de producción industrial apoyado en la ciencia y la tecnología están teniendo fuertes efectos indeseados sobre el ecosistema global (Marx 1988).

Las críticas más o menos sistemáticas a la ciencia parecen haber cedido a finales de la primera década del siglo xxi. Sin embargo, el malestar latente sobre algunos de los efectos indeseados —de naturaleza indirecta— de la ciencia sobre el entorno natural, así como la complejidad de la sociedad global, hacen que resulte fundamental promover y consolidar un estado de opinión favorable para la ciencia. Una percepción que asuma que, precisamente, el avance científico y tecnológico es elemento clave para ayudar a la Humanidad a afrontar sus grandes retos. Y a reconocer que las dimensiones científica y humanística de nuestra cultura no sólo son perfectamente compatibles, sino que conjuntamente pueden y deben contribuir a mejorar de forma sostenible las condiciones de la existencia humana. Se trata de recuperar y activar la recomendación que el filósofo y educador norteamericano, John Dewey, hiciera en el siglo pasado de utilizar la ciencia para «curar las heridas causadas por la ciencia aplicada» y, en particular, impul-

sar el desarrollo de la cultura científica, la transmisión al conjunto de la sociedad de hábitos mentales y actitudes centrales de los investigadores, como la curiosidad, la objetividad, la innovación, el debate racional, la disposición a corregir las opiniones sobre la base de argumentación y evidencia empírica (Dewey 1934). En definitiva, la tradición de racionalismo ilustrado por la que abogara incansablemente Karl R. Popper, sin duda uno de los grandes filósofos y pensadores de la segunda mitad del siglo xx.

Contribuir siquiera de manera modesta a esta gran tarea, es, en definitiva, el propósito de este libro: un libro en el que destacados protagonistas de la ciencia y las artes de nuestro tiempo, investigadores en la frontera del conocimiento, pasan revista al estado del arte y las perspectivas de las ramas del conocimiento científico y artístico más características de nuestro siglo; las que, en mayor medida, están generando avances perceptibles por el conjunto de los ciudadanos y las que se dirigen a afrontar los retos más relevantes para nuestro futuro y el de nuestros hijos: la salud, las tecnologías de la información y las de las comunicaciones, los recursos naturales, el medio ambiente y el cambio climático, la generación y la distribución más equitativa de la riqueza, sin olvidar las artes, expresión de la cultura y sensor de las inquietudes sociales de nuestro tiempo.

En BBVA nos sentimos orgullosos de sumar nuestra contribución al impulso del conocimiento y la creatividad, con la edición de este libro y, de forma más permanente, a través de los Premios Fundación BBVA Fronteras del conocimiento. Agradecemos muy sinceramente a todos y cada uno de los destacados investigadores y creadores que atendieron nuestra petición de presentar de primera mano, con el conocimiento íntimo y la *auctoritas* que confiere su destacada trayectoria en el correspondiente dominio del conocimiento, una selección de cuestiones fundamentales en sus áreas de trabajo. Espero y deseo que los lectores de este libro disfruten con su lectura, tanto como nosotros al editarlo, y se sumen al reconocimiento de los miles de investigadores que diariamente se esfuerzan por hacer avanzar nuestro conocimiento del mundo natural y social y, con ello, amplían nuestra libertad de decisión y posibilidades individuales y colectivas.

Bibliografía

- Beniger, James R. *The Control Revolution. Technological and Economic Origins of the Information Society*. Cambridge, MA-Londres, Inglaterra: Harvard University Press, 1986, 2-3.
- Ceruzzi, Paul. «An Unforeseen Revolution: Computers and Expectations, 1935-1985», en Joseph J. Corn, ed., *Imagining Tomorrow. History, Technology and the American Future*. Cambridge, MA-Londres, Inglaterra: The MIT Press, 1986, 188-201.
- Dertouzos, Michael L. «The Information Revolution. Developments and Consequences by 2000A.D.», en Heinz R. PAGELS, *Computer Culture: The Scientific, Intellectual, and Social Impact of the Computer*. Nueva York: The New York Academy of Sciences, 1984, 21.
- Dertouzos, Michael L., Richard K. Lester, Robert M. Solow y MIT Commission on Industrial Productivity. *Made in America. Regaining the Productive Edge*. Cambridge, MA-Londres, Inglaterra: The MIT Press, 1989.
- Dewey, John. «Science Must Heal Wounds Made by Applied Science.» *The Science News-Letter* vol. 25, no. 666 (Enero 13, 1934), 23-30.
- Drucker, Peter. *The Age of Discontinuity*. New York: Harper and Row, 1968.
- . «*The Age of Social Transformation*», *The Atlantic*, Noviembre 1994.
- Freeman, Chris, y Luc Soete, *The Economics of Industrial Innovation* (3rd edition). Cambridge, MA: The MIT Press, 1997.
- Harasim, Linda M., ed., *Global Networks. Computers and International Communication*. Cambridge, MA-Londres, Inglaterra: The MIT Press, 1993.
- Hastie, Trevor, Robert Tibshirani y Jerome Friedman. *The Elements of Statistical Learning. Data Mining, Inference and Prediction*. Nueva York, Berlín y Heidelberg: Springer, 2003.
- Hugues, Thomas P. *American Genesis. A Century of Invention and Technological Enthusiasm, 1870-1970* (2nd Edition). Chicago: Chicago University Press, 2004.
- Machlup, Fritz. *The Production and Distribution of Knowledge in the United States*. Princeton: Princeton University Press, 1962.
- Marx, Leo. «The Neo-Romantic Critique of Science.», en *The Pilot and the Passenger. Essays on Literature, Technology, and Culture in the United States*. Nueva York y Oxford: Oxford University Press, 1988, 160-178.
- . «The Domination of Nature and the Redefinition of Progress», en Leo Marx y Bruce Mazlish (eds.), *Progress. Fact or Illusion?* Ann Arbor: The University of Michigan Press, 1998, 201-218.
- Mowery, David, y Nathan Rosenberg. *Technology and the Pursuit of Economic Growth*. Londres: Cambridge University Press, 1989.
- Norris, Pippa. *Digital Divide. Civic Engagement, Information Poverty, and the Internet Worldwide*. Cambridge: Cambridge University Press, 2001.
- Rosenberg, Nathan. «Factors affecting the Diffusion of Technology», en *Perspectives on Technology*. Armonk, Nueva York y Londres: M. E. Sharpe, Inc., 1976, p. 197.
- Simon, Herbert. «Cohabiting the Planet with Computers», en Joseph F. Traub, ed., *Cohabiting with Computers*. Los Altos, CA: William Kaufmann, Inc., 1985, 153-172.
- . *The Sciences of the Artificial* (3rd edition). Cambridge, MA-Londres, Inglaterra: The MIT Press, 1996.
- United Nations, *The Millennium Development Goal Report*. Nueva York, 2007.
- World Bank, *World Development Report 2008*. Washington, 2008.

el lugar de la ciencia en nuestra cultura en el fin de la «era moderna»

GERARD HOLTON

Nota introductoria

La esencia misma de la democracia es que cualquier intento de imposición de medidas autoritarias por parte de una institución invita a su examen casi automático con una contra argumentación razonada. Eso también es cierto, y lo ha sido durante siglos, para la autoridad que se le ha otorgado a la ciencia y para el lugar que ocupa en nuestra cultura.

Pero en ocasiones ocurre que esas contra argumentaciones razonadas se han visto oscurecidas por una marea de ataques encendidos, irracionales e incluso sensacionalistas al lugar que debe ocupar el conocimiento científico (uno piensa aquí, por ejemplo, en el movimiento conocido como «bancarrota de la ciencia» que se produjo en el siglo XIX). Hace algunos años pareció iniciarse un proceso idéntico, justo cuando las páginas que siguen fueron escritas con el propósito de ilustrar y comprender este fenómeno social, así como para alertar a sectores de la, por lo general, plácida comunidad científica del peligro y animarles a actuar contra él.

Entonces existía la esperanza de que —en parte debido a los extraordinarios avances que constantemente se producían en la ciencia moderna y a sus aplicaciones prácticas en la vida diaria— esas voces extremistas fueran silencia-

das. No ha sido así. De hecho, una combinación de distintas fuerzas ha estado activa (al menos en Estados Unidos y algunos países europeos) para poner en marcha el péndulo del antagonismo en contra de la autoridad de la ciencia en los círculos académicos, en la cultura popular, entre políticos de gran visibilidad e incluso entre algunos teólogos. Han aparecido un número cada vez mayor de libros con títulos como El fin de la ciencia; de publicaciones especializadas cuyos argumentos centrales son que la esencia del método científico «surgió a partir de la tortura humana trasladada a la naturaleza»; de ataques altamente fundados a la biología evolutiva; un sentimiento creciente entre determinados filósofos y sociólogos posmodernos que aducen que estamos asistiendo al «fin de la modernidad» y que el concepto de «naturaleza», al carecer de validez, convierte el ejercicio de la ciencia en un mero intento de hacer carrera, y un intento de silenciar, por parte de altas esferas de gobierno, hallazgos científicos consensuados relativos a los peligros que amenazan el medioambiente y la salud pública.

En suma, las observaciones y conclusiones expuestas a continuación acerca del lugar de la ciencia de nuestra cultura cobran hoy una relevancia especial.

Detrás de cada acción en la vida de un científico —ya sea la elección de un proyecto de investigación o su interacción con los estudiantes, el público y los medios de comunicación, o la búsqueda interminable de fondos, o consultas solicitadas por representantes del gobierno— siempre hay un factor oculto que en gran medida determina el resultado. Ese factor es cómo la sociedad en general interpreta el lugar de la ciencia en la cultura. La mayoría de quienes practican la ciencia afirmarían que carecen del interés o del conocimiento suficientes como para preocuparse de un problema tan complejo y amorfo en apariencia, al menos no hasta que llega el momento —como periódicamente sucede— en el que empiezan a percatarse de que sus suposiciones en gran parte inconscientes sobre las relaciones entre ciencia y el conjunto de la sociedad están siendo gravemente puestas en tela de juicio.

Esos días han vuelto. Aquí y allá los intelectuales empiezan a enfrentarse al hecho de que, cada vez más, conceptos como «el fin de la era moderna», «el fin del progreso» y «el fin de la objetividad», procedentes de ciertos sectores académicos, de divulgadores particularmente elocuentes e incluso de miembros del Congreso se están asentando en la mentalidad popular con sorprendentemente escasa oposición por parte de los líderes de la comunidad científica. Pero, lejos de tratarse de una fase pasajera, este movimiento —diferente de la ola «anticiencia» que he analizado en otros escritos—¹ marca el resurgir de una rebelión ya antigua y recurrente contra algunos de los presupuestos de la civilización occidental que tienen su origen en la Ilustración, en especial contra la afirmación de que la ciencia puede conducir a una clase de conocimiento que es progresivamente no susceptible de ser probado, universalmente accesible (es decir, intersubjetivo) por principio y potencialmente valioso y civilizador. El impacto de este sentir colectivo en la vida del científico, en la comprensión pública de la ciencia en general y en la legislación que la regula crece cada día y resulta evidente hasta para los más despistados.

El objeto de este ensayo es ayudar a la comprensión de este movimiento, de sus principales fuentes y los propósitos que lo impulsan. Para ello empezaré con un repaso de algunos de los principales teóricos sobre la cuestión de qué papel debe desempeñar —si es que debe desempeñar alguno— la ciencia en nuestra cultura y de sus efectos en la legislación sobre este tema en Estados Unidos, que en este momento está rediseñando los objetivos y la práctica de la disciplina científica. Para ello es necesario mirar al pasado, más allá del llamado «contrato» implícito entre ciencia y sociedad fraguado en el periodo inmediatamente posterior a la Segunda Guerra Mundial.

Dicho contrato, que continúa siendo el mito dominante entre la mayoría de los científicos aunque a duras penas se corresponde con la realidad de hoy, fue el resultado de una fase de inocencia, cuando durante unas pocas décadas la búsqueda del conocimiento científico se consideraba

en líneas generales —y en especial por los científicos— la encarnación de los valores clásicos de la civilización occidental, empezando por las tres virtudes primarias de verdad, bondad y belleza; cuando la ciencia era alabada por su afán por la verdad y vista como un proceso de iluminación en la cultura moderna, una especie de búsqueda newtoniana de la omnisciencia; cuando se pensaba que la ciencia encarnaba una escala de valores positivos dentro de un mundo imperfecto, tanto a través de su tradicional práctica del comportamiento honorable como por su tendencia a desembocar en aplicaciones capaces de mejorar la condición humana y alejar a los enemigos que amenazaban nuestro modelo de sociedad: una búsqueda baconiana de una forma benigna de omnipotencia. Cuando el descubrimiento de la verdad en la estructura, coherencia, simplicidad y racionalidad del mundo era considerado, en suma, una suerte de hechizo kepleriano, la más alta recompensa a tan ardua tarea.

Antes del fin de la euforia

La última vez que la descripción optimista que acabo de hacer ha sido dada por cierta de forma generalizada, al menos en Estados Unidos, fue durante el periodo que siguió al fin de la Segunda Guerra Mundial. También estaba resumida en el famoso informe de Vannevar Bush, titulado *Science, the Endless Frontier* [Ciencia. La frontera interminable], de 1945, que se convirtió en la principal fuerza impulsora de la ciencia en este país. Puesto que resulta un ejemplo especialmente esclarecedor del optimismo posilustración acerca del papel de la ciencia en la cultura, y uno que muchos científicos siguen considerando vigente, resulta interesante examinar algunas de sus ideas principales.

En noviembre de 1944 el presidente Franklin D. Roosevelt solicitó a Vannevar Bush, director de la Oficina de Investigación y Desarrollo Científicos durante la guerra, un informe que resumiera cómo, en el mundo de la posguerra, la investigación de las ciencias de la naturaleza —él las llamó «las nuevas fronteras de la mente»— podían ser fortalecidas y puestas al servicio de la nación y de la humanidad en general. Roosevelt estaba particularmente interesado en tres objetivos: librar una «nueva guerra de la ciencia contra las enfermedades», «descubrir y desarrollar el talento científico entre los jóvenes estadounidenses» y diseñar un nuevo sistema de fuerte apoyo federal a la investigación científica en los sectores público y privado. Más allá de estos resultados concretos, aducía que las aplicaciones de la ciencia, tan útiles durante la terrible guerra para salvar al mundo de la amenaza fascista (en alusión al éxito que supusieron el radar y los mecanismos antisubmarinos de los aliados), podían ahora reconducirse «en un uso más completo y fructífero» que dotara a la ciencia de «una vida más completa y fructífera» también.

La respuesta detallada de Vannevar Bush a esta solicitud llegó menos de ocho meses después, el resultado de un programa de urgencia desarrollado por un deslum-

¹ Gerald Holton. *Science and Anti-Science*. Cambridge, Massachusetts: Harvard University Press, 1993 [Ed. esp. *Ciencia y anticiencia*, Nivola, Madrid, 2003].

brante comité de 40 expertos procedentes de la industria, el mundo académico y el gobierno. Para entonces Roosevelt había muerto, pero con el final de la guerra a la vista, la administración estadounidense se mostró especialmente receptiva a las ideas propuestas en el informe. Si bien algunos detalles eran en exceso optimistas y otros tuvieron que ser modificados en la práctica (a menudo para escándalo de Bush), es un hecho generalmente aceptado que la visión de éste sentó los cimientos para el desarrollo de nuevas instituciones en apoyo a la ciencia durante las décadas siguientes y gozó de la actitud popular favorable necesaria para llevar a cabo las acciones. Así, se sentaron las bases para un liderazgo global en muchas ramas de la ciencia básica. Hasta el recrudecimiento de la guerra de Vietnam no hubo un descontento significativo del pueblo hacia la autoridad del gobierno y también hacia la evidente aplicación de avanzadas tecnologías en una guerra inútil e impopular, y por ende hacia la ciencia que, supuestamente, había hecho posibles tales abusos. Aquello marcó el fin de lo que podría considerarse una fase de euforia en las relaciones entre ciencia y sociedad en el siglo xx.

El informe Bush, así como las propuestas rivales del senador Harley Kilgore, fueron ejemplos históricos del progresismo basado en la ciencia que reinaba en aquel momento, según el cual ciencia y democracia eran aliadas naturales al servicio del ideal de fortalecimiento e instrucción de la sociedad en su conjunto. En este sentido formaban parte del sueño americano que se remontaba a Benjamin Franklin y sus contemporáneos, hombres de Estado aficionados a la ciencia. El mismo Vannevar Bush así lo apuntaba en el breve prólogo a su informe, en el que afirmaba inspirarse para su empresa en «el espíritu de los pioneros, aún vivo en nuestra nación». Y para hacer aún más explícita la relación con la tradición encarnada por Condorcet, añadía una frase que, aunque representaba la opinión dominante de un ciudadano de mediados de la década de 1940, hoy sería rechazada por muchos que se consideran hijos de las de 1960 y 1970. Bush escribió: «El progreso científico es la llave a la seguridad nacional, a una mejor salud, a mejores empleos, a un nivel de vida más alto y al progreso cultural». En dicha afirmación resuenan los ecos de la fórmula de Thomas Jefferson: «Las verdades esenciales [son] que el conocimiento es poder, el conocimiento es seguridad, el conocimiento es felicidad».

Bush y sus contemporáneos habrían difícilmente imaginado que para principios de la década de 1990 esas creencias empezarían a ser rechazadas, incluso en las más altas esferas. Tampoco habrían concebido que, por ejemplo, un importante miembro del Congreso de Estados Unidos para política científica llegaría a sugerir (como después veremos con más detalle) que la ciencia y la tecnología son las verdaderas culpables de la triste lista de fracasos producto de décadas de liderazgo político equivocado. Dijo: «El liderazgo global en ciencia y tecnología no se ha traducido en liderazgo en salud infantil, esperanza de vida,

índices de alfabetización, igualdad de oportunidades, productividad de los trabajadores o ahorro de energía. Tampoco ha servido para mejorar nuestro maltrecho sistema educativo, la decadencia de nuestras ciudades, la degradación del entorno, el deficitario sistema de sanidad y la mayor deuda nacional de la historia».² Y otro observador de gran prestigio, en otro tiempo director de la Fundación Nacional para la Ciencia (National Science Foundation), anunciaba exultante: «Los días de Vannevar Bush se han ido para siempre... el mundo entero está cambiando».

El cambiante equilibrio de los sentimientos

Tras este breve repaso de la visión mundial predominante a mediados del siglo xx, antes de que la generación que ocupa hoy los puestos de responsabilidad llegara a escena, debemos descender del ámbito de los caprichos pasajeros para comprender mejor los mecanismos causales responsables de los cambios en el lugar asignado a la ciencia en periodos determinantes de la historia intelectual de los últimos cien años. Ya que sólo conociendo las causas generales de la variación en la ideología subyacente podremos comprender los cambios en la política científica en momentos determinados.

En este punto debemos enfrentarnos a la cuestión de si estos cambios son graduales y forman parte de una evolución necesaria, o son por el contrario tan repentinos que, como en una revolución política, el tránsito del final de una era al inicio de otra nueva se produce con solución de continuidad. Si éste es el caso, entonces ahora estaríamos atravesando un periodo de ruptura histórica, en que hemos dejado atrás la «modernidad» y vivimos rodeados de «posmodernidad». Aunque dudo de que éste sea el caso —ciertamente no resulta evidente en los contenidos de la ciencia, pero sí en los escritos actuales sobre ciencia—, existe una tendencia dentro de la historia propiamente dicha que lleva algún tiempo tratando de identificar el advenimiento de una nueva era. La periodización, la clasificación del flujo de los acontecimientos en etapas claramente diferenciadas es una herramienta de uso común, aunque se aplica más sabiamente desde la seguridad que proporciona la visión retrospectiva. Así es como nacieron capítulos de los libros de texto tales como «La edad de la razón» o «La era progresista en Estados Unidos» a finales del siglo xix.

Un ejemplo aleccionador de ese género nos lo proporcionó el historiador norteamericano Henry Adams. A comienzos del siglo xx quedó gratamente impresionado con las publicaciones del físico y químico J. Willard Gibbs de la Universidad de Yale acerca de las reglas de fase necesarias para comprender los equilibrios heterogéneos. A Adams también le fascinaba la extraña idea de algunos físicos de la época de que las reglas de fase pudieran servir, por analogía, como medio para clasificar en orden jerárquico los siguientes elementos: sólido, fluido, gas, electricidad, éter y espacio, como si constituyeran fases de una secuencia. Estimulado por estas ideas, Adams creía

² George E. Brown, Jr. citado en *Science*, 260: 73, 7 de mayo de 1993.

que también el pensamiento atravesaba distintas fases en el tiempo, cada una de las cuales representaba un periodo diferente. En su ensayo de 1909 «Las reglas de fase aplicadas a la Historia», Adams llegó a una notable conclusión acerca del fin inminente de la modernidad: «El futuro del Pensamiento —escribió—, y por tanto de la Historia, está en manos del físico, y [...] el historiador futuro debe buscar su educación en el mundo de la física matemática... [Si es necesario] los departamentos de Física tendrán que asumir solos esta tarea». La conclusión de Henry Adams pudo muy bien en su día considerarse una definición de cómo se pensaba entonces que sería la futura era posmoderna.

La formulación actual es más bien la opuesta. Cito este ejemplo —aunque me vienen a la mente otros muchos— para hacer notar la incomodidad que me producen estos intentos de dividir la historia en periodos estancos. Una noción menos rígida y más plausible sería reconocer que en cualquier tiempo y lugar, incluso durante un periodo en que una civilización parece encontrarse en una situación más o menos estable de equilibrio dinámico, coexisten varias ideologías en competición y conflicto dentro de la argamasa momentánea y heterogénea de perspectivas. Como apuntó Leszek Kolakowsky: «Es seguro que la modernidad tiene tan poco de moderna como los ataques a la modernidad [...] El choque entre lo antiguo y lo moderno es probablemente eterno y nunca nos desharemos de él, porque expresa la tensión natural entre estructura y evolución, y esta tensión parece tener raíces biológicas; podemos afirmar que es una característica inherente a la vida».³

En ocasiones es posible, en retrospectiva, identificar uno de los puntos de vista en liza como dominante por un periodo de tiempo más o menos prolongado. Pero también es probable que, a tiempo real, esto tenga dos efectos. El primero es que cada uno de los grupos en competición trabaja fervientemente por elevar su ideología a una posición donde sea aceptada como «el sentir de una época» o «el clima de opinión» que caracteriza una etapa y una región determinadas. La última y más ambiciosa también *intentará, como parte del programa de actuación, deslegitimar las reivindicaciones de sus principales rivales*. Especialmente cuando el equilibrio previo comienza a resquebrajarse, la algarabía de voces antagónicas se vuelve ensordecedora. Algunos triunfadores parciales pasan a convertirse en adalides, y hasta puede ocurrir que uno de ellos sea reconocido por un tiempo como el representante de una nueva visión del mundo o del «sentir» de la sociedad. El segundo efecto será, dentro de este constante ir y venir de fuerzas históricas cambiantes, que la tendencia innata en la humanidad a perderse en la ambición desmedida o en el sectarismo infecte a algunos de estos adalides (que ocasionalmente pueden ser también científicos). Esta tendencia a (como lo expresó Hegel) «perpetuarse ad infinitum» o simplemente a *ceder a los excesos* puede a su vez generar, como reacción, la misma clase de excesos por parte de las voces contrarias. Reconocer estos dos hechos

es, en mi opinión, central a la hora de comprender el curso de la cultura de nuestro tiempo,

En esta lucha continua, que empieza con el enfrentamiento entre Apolo y Dionisos en la antigua Grecia y se prolonga hasta nuestros días, la cuestión específica y más limitada del lugar asignado a la concepción científica del mundo siempre ha estado presente. En ocasiones este lugar ha ocupado el corazón de la visión del mundo triunfante o dominante, como he explicado antes; en otras ha estado con los perdedores, e incluso ha sido acusada de alimentar una gran variedad de pecados en contra del bien de la humanidad.

Los historiadores de las ideas han cartografiado las formas cambiantes de las tendencias generales antagónicas. Líderes políticos también han observado a veces con lucidez y aprensión cómo el equilibrio de los sentimientos prevalentes se trastocaba porque, como dijo Jefferson, «son la manera y el espíritu de un pueblo lo que permite que un república siga viva. Una degeneración de los mismos es un cáncer que pronto devora el corazón de las leyes y la constitución». Académicos de peso han documentado cómo una de las concepciones del mundo, y la postura científica que la sostenía, ganaron predominancia sobre las otras durante algunas décadas en segmentos significativos de la cultura occidental. Un ejemplo es el estudio pionero de Robert K. Merton sobre ciencia y el puritanismo del siglo xvii. También existe documentación abundante que demuestra que estos sentimientos con el tiempo desaparecieron, conforme el equilibrio general entre bonanza y crisis se inclinaba en el sentido contrario durante algunas décadas más. En cuanto a los científicos mismos, la mayoría ha prestado poca atención a esta continua marea de sentimientos, excepto para intervenir ocasionalmente asignándose el mérito de los vaivenes positivos, o para declararse víctimas de los negativos.

Hoy, este espectáculo de continuo ir y venir, tan apasionante para el estudioso, ha dejado de ser únicamente sujeto de estudio para historiadores. El equilibrio general entre los elementos en liza, y con él la actitud de sus adalides, está cambiando ante nuestros ojos. Estudiar ese proceso resulta tan fascinante y provechoso para un historiador de las ideas —que es el papel que adopto aquí— como la aparición de una supernova para un astrónomo. Pero en ambos casos el estado actual de las cosas es el producto de un proceso histórico, la última adquisición de una progresión variopinta de los acontecimientos.

Hacia un «siglo monista»

Echemos pues un vistazo a aquellos que se decían representantes del sentir de su época desde hace cien años hasta el presente, en una secuencia de ejemplos seleccionados con la intención de que sean análogos a las etapas observables en el crecimiento de una célula bajo la lente del microscopio. Nuestro primer ejemplo concierne a un acontecimiento sucedido cuando empezaba un nuevo siglo: la

³ Leszek Kolakowsky. *Modernity on Endless Trial*. Universidad de Chicago, 1960, p. 4.

Exposición Mundial de Columbia celebrada en Chicago en 1893. Esta exposición había sido concebida como la celebración triunfal del progreso social y humano en todos los campos, pero sobre todo en la industria, la ciencia y la arquitectura. Las grandes atracciones eran el Vestíbulo de la Maquinaria, el Edificio de la Electricidad, la Fuente eléctrica y las salas dedicadas a Transportes y a Minería. En el día de la inauguración, el presidente de Estados Unidos Grover Cleveland pulsó un botón que encendió un gran número de luces eléctricas y motores (las bombillas eléctricas y los motores de corriente alterna eran entonces aún relativamente nuevos). Esto causó que los miles de espectadores avanzaran al unísono presas de la excitación y muchos se desmayaron en el tumulto. Creo que podemos afirmar sin miedo a equivocarnos que pocos de entre los 37 millones de visitantes que tuvo la exposición se preocupaban entonces de los efectos negativos de la rápida industrialización. Y pocos, si es que hubo alguno, adivinaron que un siglo más tarde, en la Feria Mundial organizada en Corea del Sur, la exposición oficial de Estados Unidos, como obedeciendo a un nuevo *zeitgeist*, estuvo dedicada por entero a los detritus del mundo posindustrial, con montículos de maquinaria rota y fotografías de pútridos vertidos nucleares; o que la exposición permanente en el Smithsonian Museum of American History de Washington titulada «La ciencia en la vida en Estados Unidos» dedicaría la mayor parte de su espacio a ilustrar los riesgos de la ciencia y el supuesto descontento popular con la tecnología.

Otra indicación de hasta qué punto cambió la visión del mundo en el transcurso de un siglo es que uno de los grandes acontecimientos de la exposición de 1893 fue el espectacular Parlamento de las Religiones. Las creencias religiosas personales son, y siempre lo han sido, algo cercano a los corazones de los norteamericanos. Sin embargo, ahora nos sorprende que en un escenario dedicado a glorificar la ciencia y la industria, cientos de líderes religiosos procedentes de todas las partes del mundo se reunieran para presentar sus ideas en el transcurso de 100 sesiones repartidas en 17 días. Fue un acontecimiento de lo más coloristas, con hindúes, budistas, jainitas, judíos, protestantes, católicos, seguidores del sinto y de Zoroastro, etcétera, todos reunidos con sus vistosas túnicas «en un encuentro fraternal», en palabras de presidente del parlamento J. H. Barrows. El propósito era claro. Al igual que el de la Exposición en su conjunto, el lema del Parlamento de las Religiones tenía que ver con el progreso y la unidad armónica entre las gentes. Por tanto la Exposición, explicaba Barrows, no podía excluir la religión como no podía tampoco excluir la electricidad. La ciencia se invocaba como un aliado en el afán por alcanzar una unidad mundial, al tiempo que servía a las necesidades de la humanidad.

Uno de los más apasionados defensores de la idea de que la ciencia, la religión y la cultura son aspectos de un gran programa de unificación entre los pueblos fue uno de los organizadores del Parlamento de las Religiones, Paul Carus,

un editor hoy recordado fundamentalmente por haber acercado los escritos de Ernst Mach a los lectores estadounidenses. El título de su presentación⁴ era, nada menos que «La ciencia, una revelación religiosa». El suyo era esa clase de deísmo anticlerical y poscristiano, que en gran medida habría resultado atractivo para algunos filósofos-estadistas de cien años antes. La dignidad del individuo, pensaba Carus, sólo puede encontrarse a través del descubrimiento de la verdad, y ésa es la misión de la ciencia. Por tanto, anunciaba, «Dios nos habla a través de la ciencia». No se trataba de tener que elegir entre la Virgen María y la dinamo, sino más bien que los laboratorios eran las verdaderas catedrales y viceversa. Tal y como rezaba en la mancheta de su revista *The Open Court*, Carus estaba «entregado a la ciencia de la religión [y] a la religión de la ciencia».

Carus encarnaba un universalismo popular y partidario de la ciencia de aquella época que hoy está gravemente cuestionado tanto por la izquierda como por la derecha. Le he elegido porque la visión que pintaba del mundo constituye un buen ejemplo del entonces boyante movimiento *Monismo moderno*, que se basaba en la creencia de una «concepción unitaria del mundo». Surgió básicamente como reacción al dualismo cartesiano materia/mente y contra la multiplicidad de la experiencia del sentido común que partía de la individualidad de cada uno. El movimiento defensor del monismo tenía la enorme ambición, en palabras de Carus, «de dirigir todos los esfuerzos hacia la reforma y de regenerar por completo nuestra vida espiritual en todas sus manifestaciones». Ello implicaba por supuesto sustituir la religión convencional con lo que Carus llamaba la «Religión de la Verdad», donde la Verdad se define como «la descripción del hecho [...] comprobable de acuerdo a los métodos de la investigación científica». En este sentido, la ciencia es «revelación»; y de esta manera era posible superar el antiguo e inaceptable dualismo verdad científica/verdad religiosa.

A la cabeza de un reducido pero ambicioso movimiento monista internacional estuvo el gran químico alemán Wilhelm Ostwald (premio Nobel, 1909). Aunque la mayor parte de los científicos modernos son muy conscientes de los límites existentes incluso en su campo de investigación —como Max Planck dijo en 1931, «una ciencia nunca está en situación de resolver por completo y de manera exhaustiva el problema al que se enfrenta»— las publicaciones del movimiento monista demuestran que aspiraba a resolver todos y cada uno de los aspectos de la cultura, desde la educación de los niños hasta la economía de las naciones y, por supuesto, empleando métodos de investigación científica. Así Ernst Haeckel, otro impulsor del movimiento, predijo que las ciencias físicas con el tiempo reducirían el origen de la materia «a un único elemento».

A pesar de la ingenuidad filosófica de sus líderes, el movimiento atrajo durante un tiempo entusiastas seguidores. En Alemania tuvo representaciones en 41 ciudades e incluso organizó manifestaciones multitudinarias con-

4

John Henry Barrows (ed.). *The World's Parliament of Religions*, v. II. Chicago: The Parliament Publication Co., 1893, pp. 978-981.

tra la Iglesia. El éxito del movimiento puede explicarse en parte por el hecho de que sus seguidores habían tenido que vivir bajo el clericalismo político y reaccionario de Alemania. Pero he elegido expresamente este ejemplo de excesivo «cientificismo» por parte de una pequeña minoría de científicos como primera ilustración de *la retórica de una ambición polarizadora de muchos movimientos antes, después y en posturas contrarias*. Llevado por su fervor, Ostwald, con un orgullo desmesurado y jamás igualado por los escasos partidarios de científicismo que sobreviven en la actualidad, alcanzó las más altas cotas de ambición desmedida con reivindicaciones tales como éstas, que hizo en 1911: «De la ciencia esperamos lo más elevado que la humanidad es capaz de producir [...] Todo lo que la humanidad, en términos de deseos y esperanzas, objetivos e ideales, resume en el concepto de Dios, la ciencia lo cumple». Y, por último: «La ciencia ahora y con éxito inconmensurable ocupa el lugar de lo divino». Ostwald añadía la profecía de que «asistimos a la llegada del Siglo Monista [...] que inaugurará una nueva época para la humanidad, del mismo modo que hace 2.000 años la predicación del amor a la humanidad había inaugurado una época».⁵

Poco después de esta publicación la llamada a la bondad y al amor al prójimo que el monismo y el cristianismo promulgaban estaban lejos de haber triunfado. En su lugar lo había hecho la guerra, que William James llamó «la sanguinaria enfermera de la Historia». Por extraño que parezca, fue Henry Adams quien presintió el fin del siglo monista. En su autobiografía, escrita en 1905 y titulada *La educación de Henry Adams*, señalaba que el curso de la historia se alejaba de la Unidad y caminaba hacia la fragmentación y la multiplicidad. Ciertamente, en el periodo inmediatamente posterior a la Primera Guerra Mundial la idea de progreso y el optimismo acerca del lugar de la ciencia en la cultura habían muerto. La balanza se había inclinado hacia el lado contrario. El único movimiento de envergadura con ambición política que continuaba reivindicando una base científica era por supuesto el marxismo, en especial como lo defendía Lenin en su libro de 1908 *Materialismo y empiriocriticismo*. La afirmación de que el marxismo-leninismo, la ideología fundadora de la Unión Soviética, tiene algo que ver con la verdadera ciencia es un mecanismo retórico, uno más de las grandes falsedades de aquel periodo histórico, incluso si se trataba de propaganda enseñada a todos los niños de los países comunistas. Está demostrado que no es cierta, tanto por el análisis defectuoso que de la ciencia y de la filosofía hace el propio libro de Lenin, como por el maltrato generalizado de que fueron víctimas los científicos soviéticos cuando sus teorías no eran del agrado de sus gobiernos.

La predicción de Spengler del fin de la ciencia

Lo que probablemente constituye el ataque más difundido contra la reivindicación optimista de la ciencia se publi-

có cuando la Gran Guerra tocaba a su fin, en 1918: un libro que más tarde influyó a teóricos de la historia como Arnold Toynbee y Lewis Mumford. Se titulaba *La decadencia de Occidente* y lo escribió un profesor de matemáticas alemán llamado Oswald Spengler. Un breve resumen no puede hacer justicia a aquella obra densa y barroca, pero la cuestión que me interesa subrayar aquí es lo que el libro afirmaba acerca del tema que nos ocupa. Spengler venía a decir que en todos los sectores de la humanidad, en todas las épocas desde Egipto, Grecia y Roma, la historia de la civilización había seguido básicamente un mismo curso, y que así sería siempre. Por tanto el destino inevitable de Occidente era terminar reducido a cenizas en un plazo de tiempo que, afirmaba, él podía calcular a partir de los precedentes existentes. Spengler predijo que el fin de Occidente se produciría en el año 2000.

El ocaso de toda civilización, escribió, puede detectarse examinando sus ideas científicas: por la adopción de una noción de causalidad en lugar de destino, por la atención prestada a abstracciones tales como el espacio infinito y la relación causa-efecto en lugar de a «la naturaleza viva». La primacía del alma es reemplazada por el intelecto; las matemáticas impregnan más y más actividades y la naturaleza se reinterpreta como un entramado de leyes dentro del corpus que Spengler llama «irreligión científica». Aquí es donde introduce su idea más sorprendente, una que, revestida de matices modernos, vuelve a sernos familiar. Advierte que una de las características del periodo crepuscular de una civilización es que cuando precisamente la alta ciencia es más fructífera, las semillas de su destrucción comienzan a dar fruto. Ello se debe a dos razones: la autoridad de la ciencia fracasa dentro y fuera de sus límites como disciplina y un elemento destructivo y anticlerical surge dentro del cuerpo científico mismo y termina por engullirlo.

El fracaso de la autoridad de la ciencia fuera de los laboratorios, dice, se debe en gran medida a su tendencia a aplicar de forma errónea las técnicas de pensamiento que son apropiadas exclusivamente para el ámbito de la naturaleza. Spengler sostiene que el estilo de pensamiento propio del análisis científico, fundamentalmente «razón y cognición», fracasa en áreas en las que lo que en realidad hace falta es la «práctica de la percepción intuitiva», del tipo que él identifica con el alma apolínea y con la filosofía de Goethe. Al defender la existencia de una brecha insalvable entre una «racionalidad» pura de ciencia abstracta y la vida intuitiva tal como se vive, Spengler cae en el mismo error que todos los críticos anteriores y posteriores a él, de los cuales muy pocos parecen haber profundizado en la ciencia más allá de sus libros de texto del colegio. Por tanto ignoran la enorme diferencia entre, por un lado, la «ciencia pública», el resultado final de negociaciones intersubjetivas para llegar al menos a un consenso temporal y global basado en la experimentación y la lógica, y por otro el trabajo científico previo y «privado», en el que las preferencias intuitivas, estéticas y en definitiva

5 W. Ostwald. *Monism as the Goal of Civilization*. Hamburgo: Comité Internacional de Monismo, 1913, p. 37.

no lógicas pueden ser la clave del progreso individual más allá del nivel de la ciencia pública. La complementariedad entre estas dos fases bien diferenciadas del desarrollo real de cualquier proyecto científico explica por qué, en cualquier campo, los hallazgos de los científicos que operan en culturas y estilos muy diferentes terminan enjaezados y empaquetados como productos corrientes de validez global (aunque con fecha de caducidad).

Todo esto puede resultar evidente para científicos practicantes. Pero, continúa Spengler, incluso dentro del cosmos de la naturaleza hay un ataque a la autoridad de la ciencia, el cual surge del corazón de sus propios dominios. Cada concepción es, en su base, «antropomórfica» y cada cultura incorpora este peso en la concepción clave y en los exámenes a su propia ciencia, los cuales por tanto devienen ilusiones culturalmente condicionadas. Nuestro afán de progreso científico en el siglo xx tan sólo oculta el hecho —piensa Spengler— de que, como en el periodo clásico, la ciencia está destinada una vez más a «morir por su propia espada» y dar paso así a una «segunda religiosidad».

Lo que Spengler consideraba la orgía de dos siglos de ciencias exactas estaba tocando a su fin, junto con todo lo demás de valor en la civilización occidental. A modo de posdata, en un segundo libro titulado *El hombre y la técnica* (1931) Spengler añadió que las tecnologías avanzadas, con sus productos en constante proliferación, también terminarán por socavar los cimientos de la civilización occidental porque, profetizaba, el interés y su apoyo por la ciencia y la ingeniería disminuirán: el «metafísicamente exhausto Occidente» no seguirá avanzando en estos campos. En su lugar, las antes explotadas razas del resto del mundo, «tras haberse colocado ya a la altura de sus mentores», los superarán y «forjarán un arma contra el corazón de la civilización fáustica [occidental]». Las naciones no caucásicas alcanzarán la destreza técnica, sobresaldrán en ella y la emplearán contra sus creadores caucásicos. En suma, tal y como lo explica H. Stuart Hughes, la predicción de Spengler era que Oriente triunfaría gracias a una tecnología superior, primero en el comercio, y después en el terreno militar.⁶

Una concepción científica del mundo: el Círculo de Viena

La primera reacción al diagnóstico de Spengler fue, como cabía esperar, doble: por un lado estaba la aceptación generalizada y entusiasta, presente todavía hoy en personas que nunca han leído a Spengler pero que, por así decirlo, han *mamado* sus ideas. Por otro, entre los detractores de las predicciones de Spengler figuraron por supuesto muchos preeminentes científicos. Algunos de éstos habían formado un grupo de estudio que se denominaba a sí mismo Círculo de Viena, que se reunía en la década de 1920 y principios de la de 1930 para discutir ideas y publicarlas. Incluía a Moritz Schlick, Rudolf Carnap, Philipp Frank, Kurt Gödel y Otto Neurath. Entre sus simpatizantes se encon-

traban Hans Reichenbach y Richard von Mises en Alemania, y B. F. Skinner, P. W. Bridgman, Charles Morris y W. V. Quine en Estados Unidos.

La publicación más influyente del núcleo duro del Círculo fue un breve panfleto lanzado en octubre de 1929 a modo de manifiesto del movimiento y cuyo título principal era nada menos que *La concepción científica del mundo*,⁷ un auténtico toque de trompeta en la lucha por inclinar de nuevo la balanza, por devolver a la ciencia al centro de la cultura moderna, y en contra de lo que el texto definía, ya en su primera línea, como la alternativa principal, la tendencia hacia el pensamiento metafísico y teologizante, viejo pilar del movimiento romántico.

Aunque la mayoría de los estudiosos que participaron en el Círculo de Viena se preocupaban fundamentalmente por el estudio de cuestiones epistemológicas y lógicas presentes en los cimientos de la ciencia, subyacía en ellos una clara corriente de ambiciones culturales, sociales, políticas y pedagógicas. Porque, tal y como decía el manifiesto: «La atención a las cuestiones de la vida está más estrechamente relacionada con la concepción científica del mundo de lo que a simple vista podría parecer [...] Por ejemplo, los esfuerzos por la unificación de la humanidad hacia una reforma de la escuela y de la educación muestran todos ellos un íntimo vínculo con la concepción científica del mundo [...] Debemos recurrir a las herramientas intelectuales en la vida diaria [...] La vitalidad que se manifiesta en los esfuerzos por llevar a cabo la transformación racional del orden social y económico también está presente en una concepción científica del mundo».

Los miembros del Círculo se asociaban de forma explícita no con los platónicos o los pitagóricos, sino con los sofistas y epicúreos, «con aquellos que defienden un ser terrenal, el Aquí y el Ahora». Una ciencia libre de la metafísica sería una ciencia unificada; no se enfrentaría a enigmas irresolubles; instruiría el pensamiento de modo que pudiera discernir con claridad entre el discurso relevante e irrelevante, entre intelecto y emoción, entre las áreas de estudio científico por un lado y el mito, por otro. Del mismo modo que este enfoque clarificaría, por su formulación misma, los fundamentos de la matemática, de la física, de la biología y la psicología, también desmitificaría los fundamentos de las ciencias sociales, «y, en primer lugar [...] de la historia y la economía». Esta actitud empírica y antimetafísica ayudaría a rechazar concepciones tan peligrosas como «el espíritu popular» y nos «liberaría de prejuicios inhibidores».

Así, de los «escombros del milenio» surgiría «una imagen unificada del mundo», libre de creencias mágicas. Las luchas económicas y sociales de la época se atemperarían porque «el público en masa» rechazaría las doctrinas que lo habían conducido por el camino equivocado. Además de todo esto, el espíritu de la concepción científica del mundo penetraría «en un grado cada vez mayor en la vida pública y privada, en la educación, la crianza, la arquitectura y la configura-

⁶ H. Stuart Hughes y Oswald Spengler. *A Critical Estimate*. Nueva York: Charles Scribner's Sons, 1952.

⁷ Rudolf Carnap, Hans Hahn y Otto Neurath. *Wissenschaftliche Weltauffassung: Der Wiener Kreis*. Viena: Artur Wolf Verlag, 1929.

ción de la vida económica y social de acuerdo a principios racionales». Y el manifiesto para una nueva modernidad terminaba con la encendida afirmación, en letra cursiva: «El mundo de la ciencia sirve a la vida, y la vida lo recibe».

Tal vez la más cuidada de las publicaciones que expresaban la postura del Círculo de Viena respecto a la ciencia y el racionalismo como vía para una concepción más cuerda del mundo fue el libro de Richard von Mises, científico, matemático, ingeniero y filósofo (además de estudioso de Rainer Maria Rilke) austriaco. Von Mises tituló la que está considerada como su obra más importante, no sin cierta ironía, *Kleiner Lehrbuch des Positivismus* [Pequeño manual del positivismo].⁸ Su objetivo era no sólo demostrar en qué consistiría una concepción científica, empírica y racionalista del mundo, qué herramientas emplearía y qué problemas sería capaz de resolver dentro del ámbito de las ciencias, desde las matemáticas y la física hasta la biología y las ciencias sociales. Todo ello se hace en gran detalle; pero otro propósito igualmente poderoso del libro era presentar una alternativa a las entonces dominantes opciones en la Europa de habla germana: el kantismo en Alemania y el clericalismo metafísico en Austria, que por entonces se empezaban a teñir de ideologías totalitarias. Von Mises hacía notar su asaz explícita oposición a lo que el llamaba «negativismo», en el que incluye a los antintelectualismos sistemáticos, filosóficos y políticos que permanecen presentes en el panorama actual. Entre los ejemplos que citaba estaba de hecho Oswald Spengler y el otrora popular filósofo alemán Ludwig Klages, cuyo punto de vista resultaba evidente ya desde el mismo título de su obra principal, *La mente como enemiga del alma*.

A modo de indicación de que su objetivo principal era situar a la ciencia en el centro de una cultura saludable en el sentido más amplio del término, el libro de Von Mises hablaba largo y tendido sobre la forma en que la concepción científica del mundo iluminaría la comprensión de la metafísica, la poesía, el arte, las leyes y la ética. La univocidad subyacente en los distintos logros culturales se debía, a juicio de Von Mises, a la similitud de sus métodos, siempre que éstos se aplicaran de forma racional y prudente. Los lectores originales del libro debieron de sentirse en presencia de un seguidor moderno de Auguste Comte. La última frase es, por así decirlo, el resumen de todo el proyecto: «Del futuro esperamos que gracias al aumento continuado del conocimiento científico, es decir, al conocimiento científico formulado de una manera coherente, podremos regular la vida y la conducta del hombre».

Freud: pasiones instintivas frente a intereses razonables

Pero ahora veremos cómo el peso de los sentimientos inclina una vez más la balanza, y además en la cuestión misma de si el conocimiento formulado de manera científica puede llevar a la humanidad a una conducta más racional. En 1929, el mismo año en que se publicó el opti-

mista manifiesto del Círculo de Viena, Sigmund Freud escribió y dio a conocer en esta misma ciudad un libro producto de sus años de madurez en el que ofrecía su respuesta sombría y pesimista a idénticas cuestiones. Para el fundador del psicoanálisis, la función de la ciencia en nuestra cultura había sido una preocupación constante, y en 1911 aún había sido lo suficientemente optimista como para firmar el *Aufruf* (proclama) de la Sociedad de Filosofía Positivista. Pero en aquel libro de finales de 1929, *El malestar de la cultura*, Freud encontraba que la ciencia, aunque se contaba entre las manifestaciones más visibles de la civilización, era como mucho una influencia benigna en una lucha titánica de la cual dependía el destino de nuestra cultura. Esta lucha, afirmaba, se centraba en el —a menudo condenado al fracaso— esfuerzo humano por dominar «el instinto de agresión y autodestrucción». Freud afirmaba ver, y así lo explica en el último párrafo de su libro, que «los hombres han obtenido el control de las fuerzas de la naturaleza hasta tal extremo que, con ayuda de éstas, puede que no les resulte difícil exterminarse los unos a los otros hasta que no quede ninguno».

Freud sostenía que las restricciones que la humanidad impone a las exigencias de nuestros instintos generan un antagonismo irremediable entre dichas restricciones y el «instinto destructivo» o «instinto de muerte» innatos, la tendencia que está permanentemente en lucha con el proyecto civilizador para elevar la condición moral del hombre. Escribió «[...] el instinto agresivo propio del hombre, la hostilidad que siente un hombre hacia todos los demás y la que sienten todos contra todos se opone a este programa civilizador. Este instinto agresivo es consecuencia y principal representante del instinto de muerte que encontramos junto al de Eros, y que comparte con este último el control sobre el mundo. Y ahora, creo, el significado de la evolución de la civilización ya no nos resulta tan oscuro. Debe contener la lucha entre Eros y la Muerte, entre el instinto de vida (*Lebenstrieb*) y el de destrucción (*Destruktionstrieb*), tal y como se da en la especie humana. Esta lucha es la esencia de la vida, y la evolución de la civilización puede por tanto describirse sencillamente como la lucha de la especie humana por vivir. Y es una batalla contra molinos de viento, esa que nuestras niñeras tratan de dulcificar cantándonos nanas que hablan de la gloria celestial».

En este conflicto, la actividad científica y otras manifestaciones culturales resultan de una satisfactoria, si bien incompleta, «sublimación de las metas instintivas» que convierten a la ciencia a primera vista en una «vicisitud impuesta por la fuerza a los instintos de la civilización». Los logros de la ciencia y la tecnología se originaron y fueron saludados como herramientas que contribuirían a proteger al hombre frente a las fuerzas hostiles de la naturaleza; hoy se han tornado «adquisiciones culturales» que «no sólo suenan a cuento de hadas, sino que son, de hecho, la realización de todos —o casi todos— los deseos de un cuento de hadas» y casan con nuestras ideas de

8 Para la edición en inglés accedió a un título más sencillo: *Positivism: A Study in Human Understanding*. Harvard University Press, 1951.

«omnipotencia y omnisciencia». El hombre «se ha, por así decirlo, convertido en una suerte de Dios prostético».

Pero he aquí el problema. La felicidad está todavía lejos. «El hombre de hoy no se siente feliz dentro de su personaje conformado a imagen y semejanza de Dios», ya sea individualmente o en grupo. Ello, una vez más, se debe a que la «civilización está construida sobre la renuncia de los instintos», tales como la sexualidad o la agresividad, y «presupone precisamente la no satisfacción (a través de la supresión, la represión u otros medios) de instintos poderosos». De ahí la frustración cultural (*Unbehagen*), que domina por completo el ámbito de las relaciones sociales entre seres humanos.

La pesimista conclusión de Freud es la siguiente: «Como consecuencia de esta hostilidad mutua primaria entre seres humanos, la sociedad civilizada vive bajo la amenaza perpetua de desintegración. El interés del trabajo en común no basta para mantenerla unida; las pasiones instintivas son más fuertes que los intereses razonables [...] A pesar de todos los esfuerzos, los intentos de la civilización han sido prácticamente vanos [...] Siempre es posible unir a un número considerable de gente en la amistad, mientras que queden personas suficientes para recibir las manifestaciones de su agresividad», como ocurre en las persecuciones religiosas o étnicas.

Durante las décadas transcurridas desde que esto se escribió la historia moderna ha resultado ser, en demasiadas ocasiones, la prueba tangible de las sombrías teorías de Freud, según las cuales ni la ciencia ni ninguna otra actividad cultural pueden desplazar por completo nuestra naturaleza animal de su posición central, y sólo pueden retrasar el fin último que se cierne sobre nosotros.

Los científicos como «traidores a la verdad»

Examinemos ahora un periodo más reciente. Estamos familiarizados con las fluctuaciones de opinión entre expertos y el público en general respecto a las interacciones entre ciencia y sociedad durante las décadas de 1960 y 1970. Pero a principios de la de 1980 entró en el debate un nuevo y poderoso elemento que es objeto cada vez más de atención e institucionalización, al menos en Estados Unidos. Este nuevo elemento, esta nueva fuerza que contribuye a minar la credibilidad de la ciencia es la insistencia por parte de algunos sectores —aunque cada vez recluta más adeptos entre la población general— de que la búsqueda de la ciencia es, y ha sido siempre, desde los días de Hiparco y Ptolomeo, algo esencialmente corrupto y retorcido. En consecuencia deben aplicarse severas medidas a la práctica de la ciencia procedente del exterior. Esta afirmación, que se ha hecho oír más y más durante los últimos años en libros, informes oficiales y cientos de artículos, se ha extendido a lecturas dramatizadas, a determinadas agencias gubernamentales, a la burocracia universitaria y a distintas profesiones. La salvaguarda de la práctica ética y de los usos de la ciencia, que goza de una larga tra-

dición dentro de la comunidad científica, debe ponerse ahora, parece ser, en manos mejores y más preparadas.

Un ejemplo sorprendente y revelador de esta afirmación fue el libro escrito por dos editores científicos neoyorquinos de gran influencia, William Broad y Nicholas Wade. Su intención está ya explícita en la cubierta: *Betrayers of de Truth: Fraud and Deceit in the Halls of Science*⁹ [Traidores de la verdad. Fraude y engaño en la ciencia] y continúa con un verdadero bombazo ya en la primera línea: «Este es un libro sobre cómo la ciencia funciona en realidad». Más allá de la necesidad de poner en evidencia las relativamente escasas manzanas podridas presentes en cualquier cesto, algo que la propia comunidad científica reconoce como necesario por puros motivos de salud, el hecho es que este tipo de retórica se ha convertido en algo común. Tal y como proclaman este libro, y sus muchos seguidores, los relativamente pocos y tristes casos de mala práctica científica real o supuesta son la excepción que confirma la regla. El fraude y el engaño se describen como algo inherente a la propia estructura de la investigación científica.

De manera similar, el informe al Congreso realizado por el Servicio de Investigación del Congreso titulado «Mala práctica en la academia científica» afirmaba que, en grado cada vez mayor, «la ausencia de evidencia empírica que indica claramente que la mala práctica científica no constituye un problema [...] sugiere que sí es una posibilidad». Entre todos los sospechosos de dañar a nuestra república con sus malas prácticas, la ciencia es culpable mientras no se demuestre lo contrario. Es más, la tendencia ha sido recientemente a incluir en la acusación de mala práctica *científica* no sólo la falsificación de datos, el plagio y prácticas similares, también el catálogo de fechorías comúnmente asociadas a la humanidad en general, como «el uso inapropiado de los recursos económicos de las universidades, acoso sexual, discriminación racial, etcétera, consideradas ahora casi delitos comunes».¹⁰

Por su parte, la Oficina de Supervisión de Integridad Científica (OSIR por sus siglas en inglés) perteneciente al Departamento de Salud y Servicios Humanos de Estados Unidos incluyó en su definición de «mala práctica» en la ciencia, aparte de la falsificación y el plagio, aquellas «prácticas que se desvíen de forma sustancial de las comúnmente aceptadas dentro de la comunidad científica» (Código Federal: 42 C.F.R. 50.102). La intención aquí puede haber sido imitar la definición que ofreció el Tribunal Supremo de lo que constituye «obscenidad», valorando la mala práctica en relación con los estándares de conducta de la comunidad.* Sin embargo, cuando se trata del progreso científico, algunas prácticas contrarias a las en ese momento mayoritarias han resultado llevar el sello de las verdaderas innovaciones: desde introducir las matemáticas en la física en el siglo xvii pasando por la introducción de la mecánica cuántica (que tantos quebraderos de cabeza trajo a su originador, Max Planck) a la más

9
Simon and Schuster, Nueva York, 1982.

10
Science, vol. 6: 1203, 1993.

*
[Nota de la Traductora]
El Tribunal Supremo de Estados Unidos determina que para que algo sea obsceno tienen que cumplir estos tres requisitos: 1) debe ser de naturaleza lasciva, 2) debe carecer por completo de valor científico, político, educativo o social y 3) debe ir en contra de los estándares de conducta de la comunidad.

11

Tal y como lo relató el *Washington Post* el 20 de marzo de 1992.

12

Nature, vol. 367: 6 de enero de 1994. A diferencia de la mayoría de las publicaciones científicas de Estados Unidos, *Nature* ha permanecido alerta a las posibles consecuencias negativas de este desequilibrio. Véase por ejemplo la editorial de John Maddox del 17 de marzo de 1994, vol. 368: 185. Es de señalar que otras de las pocas voces que han denunciado la creciente marea de acusaciones gratuitas es también una periodista con formación científica: Barbara J. Culliton, en su ensayo: «The Wrong Way to Handle Fraud in Science», (Cómo no se debe tratar el fraude en la ciencia), *Cosmos*, 1994, pp. 34-35. Para un análisis de los costes que puede suponer para la ciencia esta incredulidad generalizada véase Steven Shapin, «Truth, Honesty, and Authority of Science» en el informe de la National Academy of Sciences titulado *Society's Choices: Social en Ethical Decision Making in Biomedicine* (Las opciones de la sociedad. La toma de decisiones sociales y éticas en biomedicina), National Academy Press, Washington D. C., 1994.

13

Estos datos me los proporcionó amablemente Donald B. Lindberg, director de la National Library of Medicine. Estos casos están muy lejos de la práctica perfectamente lícita de científicos que publican enmiendas a sus artículos cuando consideran apropiado hacer públicos ciertos errores cometidos. Eugene Garfield, en su artículo «How to Avoid Spreading Error», *The Scientist* 1:9, 1987, informa de que «de los 10.000 de artículos indexados en el SCI (Science Citation Index) desde su creación, algo más de 50.000 estaban clasificados como correcciones explícitas... Éstas van desde la simple subsanación de erratas hasta retractaciones e incluso disculpas por incluir datos "malos" o datos que no pueden ser verificados». Estas rectificaciones voluntarias suponen una tasa del 0,5%.

14

Además, la institución encargada de velar por la integridad de la investigación científica, la Office of Research Integrity of the U.S. Public Health Service (PHS), anunció en 1993 que, desde 1992 había encontrado un total de 14 investigadores culpables de mala

reciente innovación del trabajo en equipo. La definición de mala práctica que proponía la OSIR es un ejemplo más de la brecha existente entre la cultura científica y la cultura fuera del laboratorio. Deberíamos añadir aquí que, para ser justos, la directora del National Institutes of Health intervino en ese punto, objetando que una ampliación así de las prácticas consideradas delictivas «implicaría incluso al inventor de la penicilina, quien descubrió por causalidad las propiedades terapéuticas de las bacterias que crecían en un recipiente de laboratorio contaminado».¹¹

El poder de las acusaciones generalizadas contra la conducta científica tiene dos componentes. El primero es la pasmosa afirmación de que gran número de investigadores en ciencia básica son intencionadamente traidores a su misión fundamental, a saber, la búsqueda de verdades. En otras palabras, que no sólo hay unas pocas manzanas podridas, sino que todo el cesto lo está.

Sin embargo, aun con los ocasionales escándalos protagonizados por un número relativamente pequeño de investigadores científicos de los millones que hay en el mundo, esta acusación generalizada de fraude y engaño como algo inherente a la ciencia ha sido tomada tan en serio que periódicos, cursos universitarios, cursos de formación de científicos y médicos, comisiones, comités, sociedades científicas, etcétera, en Estados Unidos están intensamente preocupados y dedicados a la institucionalización de la prevención de la mala práctica científica. Los innumerables relatos de incidentes específicos, muchos de ellos escandalosos per se, pero la mayoría manipulados por el sensacionalismo, han creado en parte del público y en algunos legisladores la sensación de que los laboratorios académicos han caído víctimas de una gran plaga de deshonestidad. Tal y como apuntaba con osadía la revista *Nature*, la tendencia actual está resultando en «una lenta —y apoyada por Hollywood— erosión de la imagen pública [de los científicos] [...] [y está siendo sustituida] en la imaginación popular por la de un artista del engaño que sólo busca ganar dinero».¹² La revista *Time* contribuyó con un artículo sobre científicos que empezaba diciendo: «Los científicos, parece, se están convirtiendo en los nuevos villanos de la sociedad occidental». Una oleada de libros superventas aportaron nuevas acusaciones en forma de diatribas que tenían por objeto, en palabras de Bryan Appleyard en su polémico libro *Understanding the Present: Science and the Soul of Man* [Comprender el presente. La ciencia y el alma humana] que la ciencia necesita «una lección de humildad». Por lo que parece, nos encontramos aupados a hombros de enanos.

Lo que esta auténtica avalancha de euforia, y también las generalmente pobres y autoflageladoras respuestas a ella por parte de las instituciones científicas están impidiendo es que se lleve a cabo una rigurosa indagación de la tasa real de mala práctica entre los científicos, una verdadera investigación empírica que arroje una estimación razonable de la probabilidad de tales incidentes. Hasta

donde sé, sólo se han emprendido algunas acciones dispersas en esta dirección, pero incluso éstas sugieren que la tasa real de mala práctica (real, es decir, no supuesta, o «percibida sin pruebas concluyentes») es notablemente *baja*. Entre los datos obtenidos está por ejemplo que en la Biblioteca Nacional de Medicina, durante el periodo que va de 1977 a 1986, de los 2.780.000 artículos publicados en revistas biomédicas, sólo 41 hubieron que ser retirados porque incluían datos fraudulentos o falsificados (lo que está muy por debajo del 1% del total de las publicaciones científicas por década).¹³

Otros datos vienen a confirmar este punto. Así, la Food and Drug Administration, organismo encargado de regular la industria alimentaria y farmacológica de Estados Unidos, en respuesta a acusaciones o indicios de mala práctica en investigaciones científicas con fármacos en experimentación remitió 20 casos de sospecha de fraude y otras violaciones de la ley a la oficina del Fiscal General, que resultaron en 13 condenas a investigadores clínicos, una media de uno por año.¹⁴

Cierto. No debería tolerarse ni un solo caso. Pero incluso si la tasa actual de mala práctica fuera cien veces mayor que la que indican estas cifras, las preguntas más interesantes desde el punto de vista intelectual serían: en primer lugar, ¿por qué la ciencia en general continúa avanzando a pesar de ser la obra de seres humanos?; en segundo: ¿cómo de reducido es este índice de mala práctica en el campo de la ciencia si se compara con el de otros, desde el mundo de las finanzas y las leyes, hasta la industria, el periodismo y, por supuesto, la política? Y en tercer lugar: ¿por qué los pocos aunque ampliamente publicitados casos de mala práctica pueden hasta tal punto socavar la confianza del público y sus representantes políticos en la integridad de la investigación en general?

La ciencia como mito

La respuesta a estas preguntas es en gran parte que existe, además, otra razón para el éxito generalizado de los ataques a la credibilidad de la investigación científica. Esta segunda línea de asalto la abrió un grupo misceláneo de filósofos de la ciencia contemporánea y otros humanistas, algunos de ellos pertenecientes al llamado «núcleo duro» de la sociología constructivista, una pequeña porción de los medios de comunicación y un también pequeño pero creciente número de representantes del gobierno y aspirantes a políticos, así como de un segmento de críticos literarios y analistas políticos asociados con el movimiento vanguardista del posmodernismo. Se trata de un colectivo poderoso y elocuente del tipo de los que, en el pasado, han desafiado con éxito las visiones del mundo de su tiempo.

En líneas generales, el mensaje que pretenden transmitir no está ya basado únicamente en historias de comportamiento inaceptable por parte de algunos científicos. La acusación se ha generalizado y es ahora más seria. Por expresarlo crudamente, lo que afirman es que cier-

práctica de entre los 55.000 que recibían subvenciones anuales del PHS. (Comunicación privada de Lyle W. Bivens, director en funciones de la ORI, 20 de julio de 1993). Algunos de estos casos de mala práctica se remontaban a 1977. Para comprender mejor la escasa importancia de las acusaciones y su naturaleza por lo general pedestre, véase el informe: «Office of Research Integrity, Biennial Report 1991-1992», septiembre de 1993. U.S. Department of Health and Human Services. Para hacerse una idea de la vasta complejidad, de los costes y del trabajo así como de lo frágil del proceso de formular una acusación de mala práctica científica véase, por ejemplo, el documento de 63 páginas, disponible en el U.S. Department of Health and Human Services titulado «Departmental Appeals Board. Research Integrity Adjudications Panel. Subject: Dr. Rameshwar Sharma», Número de documento A-93-50, decisión núm. 1431 del 6 de agosto de 1993.

15
Para un estudio imparcial y bien documentado sobre los variados intereses de los sociólogos de la ciencia, véase Harriet Zuckermann, «The Sociology of Science», en Neil J. Smelser (ed.). *Handbook of Sociology*. Beverly Hills, California: Sage Publications, 1998, pp. 511-474.

16
Para un análisis serio de este asunto véase John R. Searle, «Rationalism and Realism. What is at Stake?». *Daedalus*, vol. 122, núm. 4: 55-83, 1993. Otro libro que busca dar respuesta a las distintas clases de críticas es: Paul R. Gross y Norman Levitt, *Higher Superstition: The Academic Left and its Quarrels with Science*. Baltimore, Maryland: The John Hopkins Press, 1994. Otra lectura estimulante es Frank B. Farrell, *Subjectivity, Realism and Postmodernism*. Nueva York: Cambridge University Press, 1994.

17
Isaiah Berlin, *The Crooked Timber of Humanity*, capítulos sobre Historia de las Ideas, Nueva York: Nueva York, 1992 [ed. esp. *El fuste torcido de la humanidad*. Barcelona: Península, 2002].

18
Alan Beyerchen, *Scientists under Hitler: Politics and the Physics Community in the Third Reich*. New Haven, Connecticut: Yale University Press, 1977.

tos miembros de la comunidad científica son culpables de defender la *existencia de verdades científicas*. Así pues, en realidad no hay en la ciencia nada susceptible de ser falsificado o traicionado; la ciencia es en sí misma incorregible, incluso sin la existencia de mala práctica.

Desde ese punto de vista, el negocio de la ciencia es principalmente interesado. Un ejemplo sería la creación y puesta en marcha de costosas instituciones que dicen estar dedicadas a la búsqueda de información susceptible de ser medida en cuantos y bosones y que en realidad no son más que «constructos sociales». Frente el realismo ingenuo que la mayoría de los científicos siguen practicando, y al agnosticismo de los más sofisticados, estos nuevos críticos proponen una solución radical: tal y como lo expresó un sociólogo de la ciencia: «La naturaleza no existe; tan sólo una red de comunicaciones entre científicos». La literatura académica está ahora llena de afirmaciones del tipo: «La ciencia es un mito útil» o «debemos abolir la distinción entre ciencia y ficción», o también «la ciencia es otra forma de hacer política».¹⁵

Los científicos han tendido a adoptar la noción bacciana de que la adquisición de conocimientos básicos sobre las causas e interrelaciones entre fenómenos —debidas a procesos que no son fácilmente predecibles ni comprendidos— puede darnos control sobre aquellas fuerzas de la naturaleza responsables de nuestros males. Pero ahora, el nuevo consenso nos dice que la flecha va precisamente en sentido contrario: es decir, no desde el conocimiento hacia el control, sino desde el control hacia el conocimiento, y a un conocimiento en gran medida cuestionable. Los intentos por encontrar un conocimiento de la realidad práctico y susceptible de ser compartido —mediante el uso de las facultades racionales e intuitivas de científicos individuales y por medio de sus intentos escépticos aunque colaborativos por alcanzar algún tipo de consenso— no sólo estaban condenados a fracasar, sino que, irónicamente, han conducido a los grandes desastres que han marcado el siglo xx. La era moderna, publicitada bajo la bandera del progreso, sólo ha conducido a la tragedia. El exceso de optimismo de un Herbert Spencer o un Friedrich Engels no podrán ser nunca reemplazados por una concepción más cauta del progreso, ya que el progreso no es más que una ilusión. El proyecto globalizador de la ciencia —encontrar unidades básicas y la armonía que trasciendan la variedad y la discordancia aparentes en la naturaleza— se considera contrario al impulso posmoderno que celebra la variedad individual y el derecho a la igualdad de todo estilo y manifestación posibles, de cada grupo, de cada conjunto de intereses. Asistimos, pues, al fin de la búsqueda de las causas últimas, al «Fin de la era moderna». Estamos sumidos en una «crisis de objetividad», la frase de moda para títulos de conferencias y documentos oficiales que examinaremos en breve.

Juntos, estos eslóganes del nuevo sentimiento emergente indican que el objetivo no es simplemente llamar a

la mejora de la práctica científica o a una mayor responsabilidad en su ejercicio, algo que resulta apropiado y de hecho se está llevando a cabo con gran entusiasmo. En el fondo, la meta es, para la rama principal de este movimiento crítico, la deslegitimación de la ciencia como una de las fuerzas intelectuales válidas y la remodelación del equilibrio cultural, como veremos enseguida en mayor detalle. A este respecto existe una gran diferencia en comparación con los movimientos históricos de protesta *interna*, tales como el positivismo en filosofía, los impresionistas o dadaístas en el terreno del arte, los compositores modernos en el de la música, etcétera. Aquí la fuerza impulsora no es la renovación desde dentro, sino la imposición de políticas culturales radicales venidas desde fuera.¹⁶

El desafío del movimiento romántico

Aquí nos encontramos ante un hecho esclarecedor: el enfrentamiento que presentamos no es algo nuevo, sino heredero de fuerzas históricas poderosas y duraderas. Por tanto será de utilidad repasar alguna de las fases en este sorprendente desarrollo de la nueva actitud ante la ciencia, para que así nos resulte más fácil extrapolar y prefigurar lo que nos depara el futuro. Aunque por razones de espacio sólo puedo incluir aquí algunos de los hitos más recientes, buscaré documentación en los escritos recientes de algunos de los más distinguidos pensadores antes que en, digamos, los representantes de la corriente dionisiaca.

Nuestro primer guía será Isaiah Berlin, ampliamente considerado como un historiador intelectual especialmente sensible y humano. Su colección de ensayos, publicada como quinto volumen de sus escritos reunidos,¹⁷ abre con una sorprendente dicotomía. Berlin escribe: «Hay, en mi opinión, dos factores que, por encima de todos los demás, han moldeado la historia de la humanidad en este siglo [xx]. Uno es el desarrollo de las ciencias de la naturaleza y la tecnología, sin duda el gran éxito de nuestra era y al que se ha prestado gran y creciente atención desde todos los ámbitos. El otro lo constituyen, sin duda, las grandes tormentas ideológicas que han alterado las vidas de prácticamente todos los seres humanos: la revolución rusa y sus consecuencias: las tiranías totalitarias tanto de derecha como de izquierda y la eclosión del nacionalismo y el racismo así como, en ciertos lugares, de la intolerancia religiosa, que, interesadamente, ninguno de los más perceptivos pensadores del siglo xix supo predecir».¹⁸ A continuación añade que si la humanidad sobrevive, en el transcurso de dos o tres siglos estos dos fenómenos se habrán convertido en las dos características principales de nuestro siglo, las que mayor grado de análisis y explicación requieran.

¿Qué puede pretender el autor yuxtaponiendo estos dos «grandes movimientos»? De entrada uno se siente tentado a ver una conexión en el hecho de que durante la Segunda Guerra Mundial el ingenio y el trabajo frenético de científicos al servicio de las tropas aliadas hizo posi-

ble poner fin a la tiranía, que podía de otro modo haber triunfado sobre las democracias y extenderse, al menos, en gran parte de Europa.

Pero una explicación así está fuera de lugar aquí. Lo que Isaiah Berlin tiene en mente es bien distinto. Conforme avanza en su sutil y elocuente análisis, el lector se percata de que la ciencia y la tiranía, los dos movimientos en principio diametralmente opuestos que, según Berlin, han definido y perfilado la historia del siglo xx, están de alguna manera interrelacionados, y que el desarrollo de las ciencias naturales modernas y de la tecnología ha podido, *mediante las reacciones en su contra*, contribuir de forma indirecta y no intencionada al ascenso de dichas «tiranías totalitarias».

Esta pasmosa asociación no está, por supuesto, expresada de forma explícita. Pero conforme avanzamos en la lectura del libro podemos atisbar el argumento implícito dentro del capítulo de título tan significativo como «La apoteosis de la voluntad romántica: la rebelión contra el mito de un mundo ideal». Aquí Berlin resume la cronología de algunos de los conceptos y categorías básicas del mundo occidental, en concreto los cambios en «los valores, ideales y metas seculares». Lo que le llama la atención es el abandono de la creencia en «el núcleo central de la tradición intelectual [...] desde Platón» y su sustitución por «una rebelión profunda y radical contra la tradición central del pensamiento occidental», una rebelión que en épocas recientes ha intentado forzar la conciencia occidental a emprender una nueva senda.

El núcleo central del antiguo sistema de creencias, que perduró hasta el siglo xx, estaba basado en tres dogmas que el autor resumía sucintamente de la manera que sigue: el primero es que «para todas las preguntas auténticas sólo existe una respuesta verdadera y las otras son falsas, y esto es aplicable también a cuestiones de comportamiento y forma de sentir, a cuestiones de teoría y observación, a cuestiones de valor tanto como a cuestiones de hecho». El segundo dogma es que «las verdaderas respuestas no pueden chocar unas con las otras». No pueden ser inconmensurables, sino que «deben formar un todo armónico», cuya integridad venga avalada por la lógica interna o por la total compatibilidad entre sus componentes.

Las religiones institucionalizadas y las ciencias se desarrollaron hasta adoptar su forma actual a partir de estos tres antiguos dogmas (aunque aquí cabría añadir que los científicos modernos, en la práctica, se han vuelto conscientes de la necesidad de proceder de manera antidogmática, es decir por el método de conjetura, experimentación, refutación y ensayo de probabilidades). En su estado puro, estos sistemas son en principio utópicos, puesto que están imbuidos de la creencia optimista, inherente a sus dogmas, de que «una vida formada de acuerdo a las respuestas verdaderas constituiría la sociedad ideal, la era dorada». Todas las utopías, nos recuerda Isaiah Berlin, están «basadas en la capacidad de llegar a conclusiones verdaderas

susceptibles de ser descubiertas y armonizadas, que sean ciertas para todos los hombres y en cualquier momento y lugar». Y, por extensión, lo mismo puede decirse del progreso científico y técnico, que forma parte de nuestro viaje a lo que él llama «la solución total: es decir, que a su debido tiempo, ya sea por obra de Dios o por los esfuerzos del hombre, el reino de la irracionalidad, la injusticia y la miseria tocarán a su fin; el hombre será liberado y dejará de ser un juguete a merced de fuerzas que escapan a su control [tales como] la naturaleza salvaje». He aquí una creencia que comparten los epicúreos, Marx, Bacon y Condorcet, el Manifiesto Comunista, los tecnócratas modernos y los «buscadores de sociedades alternativas».

Sin embargo, nos explica a continuación Isaiah Berlin, este importante componente de la visión del mundo moderno es precisamente lo que rechazaba el movimiento de rebelión de dos siglos de duración y que hoy se conoce como Romanticismo. Desde sus inicios en el movimiento alemán *Sturm und Drang* (tormenta e ímpetu) a finales del siglo xviii, se propagó rápidamente por la cultura occidental, con su promesa de sustituir los ideales del programa optimista basado en el racionalismo y la objetividad por la «entronización de la voluntad del individuo o de las clases sociales; [con] el rechazo de la razón y el orden en tanto que carceleros del espíritu».

Mi ejemplo favorito de la desvalorización de la ciencia promovida por la literatura del siglo xix es el antihéroe de la apasionante novela de Iván Turguénev *Padres e hijos*. Una de las grandes figuras de la literatura rusa junto con Gogol, Dostoievski y Tolstói, Turguénev fue un poeta enmarcado en gran medida en la tradición romántica del siglo xix inspirada en Goethe, Schiller y Byron, entre otros. *Padres e hijos* se publicó en 1861. Su personaje principal es Evguéni Vasílievich Bazárov, un universitario que estudia ciencias naturales y que espera sacarse pronto el título de médico. Al ser un científico «que lo examina todo desde el punto de vista crítico», se confiesa a sí mismo ser, ideológica y políticamente, un nihilista, consecuencia natural de no reconocer ningún tipo de autoridad externa. Toda conversación sobre el amor, o «sobre la relación mística entre un hombre y una mujer» es para él sólo «romanticismo, tontería, cursilería, arte». Más valdría estudiar el comportamiento de los escarabajos. Incluso cuando se marcha de vacaciones se lleva consigo el microscopio y pasa «horas y horas» sentado ante él. Leer a Puschkin, dice, es cosa de niños pequeños y opina que sería mucho mejor empezar con *Fuerza y materia*, de Ludwig Büchner, un libro publicado en 1855 y que defiende una visión tan materialista del mundo que su autor se vio obligado a dimitir de su plaza de profesor en Alemania (es, como luego se vio, un libro que Albert Einstein anotó en sus notas autobiográficas como uno de los dos o tres que más le habían impresionado de niño y que le inspiró dedicarse a la ciencia).

Lo que importa, dice Bazárov, «es que dos y dos son cuatro, y todo lo demás son tonterías». Cuando conoce a

una mujer hermosa e inteligente, deja perplejo a su amigo diciéndole que sería un cuerpo hermoso para examinar... en la mesa de disección. Pero ¡ay! el destino se venga de él y termina conduciéndole hasta el lecho de muerte de un aldeano enfermo de tifus, y se ve obligado a colaborar en su autopsia. Pero se corta con el escalpelo y pronto se encuentra al borde de la muerte, víctima de una infección quirúrgica. Mientras agoniza trata de aferrarse a la realidad preguntándose en voz alta: «¿Cuánto es 8 menos 10?». En suma, es un personaje caricaturesco y además recurrente en la literatura, excepto que en otros casos de científicos emocionalmente disfuncionales, desde el doctor Frankenstein hasta Stragelove, éstos provocan sepsis quirúrgica no sólo a sí mismos, sino también a quienes los rodean.

Volviendo a Isaiah Berlin, resulta curioso que, como él señala, nadie predijera que la visión del mundo que trajo consigo el Romanticismo sería la dominante «durante el último tercio del siglo xx». Para los «rebeldes» de nuestro tiempo, la búsqueda ilustrada de la generalización y el orden racional sólo tuvo como resultado la patética figura del científico encarnada por Bazárov, y, por tanto, debe reemplazarse por la celebración del individuo, por un antirracionalismo radical, por «la resistencia a las fuerzas externas, ya sean sociales o naturales». En palabras de Johann Gottfried von Herder, el rebelde grita: «¡No estoy aquí para pensar, sino para ser, para sentir, para vivir!». La verdad, la autoridad y la nobleza se alcanzan a través del sufrimiento heroico.

La afirmación de la voluntad individual sobre la razón colectiva ha minado lo que Isaiah Berlin considera los tres pilares básicos de la civilización occidental. La revolución romántica, por supuesto, también nos ha legado obras de arte intemporales en el terreno del arte, la música y la literatura. Pero fue el origen, por así decirlo, de un reflejo antitético, surgido como reacción a la concepción el mundo ilustrada. En la apoteosis de la voluntad romántica de nuestra época dicho reflejo se ha convertido en la alternativa, en «la autoafirmación romántica, el nacionalismo, el culto a los héroes y a los líderes y, por extensión... en el fascismo, el irracionalismo brutal y la opresión a las minorías». Además, en ausencia de «reglas objetivas», las nuevas reglas son aquellas que los propios rebeldes formulan: «Las metas no son [...] valores objetivos [...] las metas no se descubren sino que se construyen; no se encuentran, se crean».

Como resultado de ello «esta guerra en contra del mundo objetivo, en contra de la noción misma de objetividad» propuesta por filósofos pero también en novelas y en obras de teatro, infectó la concepción del mundo moderna. Los «románticos han asestado un golpe mortal» a las certidumbres del pasado, y han «alterado de forma permanente la fe en lo universal, en la verdad objetiva que determina la conducta humana». Y —podría haber añadido— también la ciencia. Como toda revuelta, ésta nos enfrenta a elecciones que parecen mutuamente excluyen-

tes. Al igual que en los casos de excesos éticamente cuestionables, como el de Ostwald, se trata una vez más de la disyuntiva y/o, antes que de la necesaria complementariedad entre las funciones racionales, pasionales y espirituales de la humanidad. Uno no puede evitar recordar el hecho de que los extremos tienden a encontrarse. Así el poeta William Blake, epítome de la rebelión romántica que calificó la obra de Bacon, Newton y Locke de «satánica», compuso en su *El matrimonio del cielo y el infierno* (1790) uno de los «Proverbios» que reflejan el credo de muchos de los actores en liza en esta historia hasta el día de hoy: «El camino del exceso conduce al palacio de la sabiduría».

La rebelión romántica impregna la política

Otros autores han trabajado para verificar y elaborar las implicaciones de los descubrimientos de Isaiah Berlin, y en especial el ominoso encuentro entre los extremos de la revolución romántica y las doctrinas políticas de irracionalidad. Esto se hizo evidente en la Revolución Cultural del la China maoísta, en la URSS y en otros regímenes totalitarios. En un intento por documentar al menos un ejemplo revelador, el historiador Fritz Stern ha estudiado las fases tempranas de la expansión de nazismo en Alemania cuando en la década de 1920 surgieron los que él llama «ludditas culturales, quienes, resentidos contra la modernidad, buscaban hacer pedazos la maquinaria de la cultura en su totalidad». La furia que les provocaba una parte esencial del programa de la modernidad, «el creciente poder del liberalismo y el laicismo», se extendió de forma natural también contra la ciencia. Julius Langbehn fue uno de los ideólogos alemanes más leídos en la década de 1920, y Stern escribe de él: «El odio a la ciencia impregnaba todo su pensamiento [...] Para Langbehn, la ciencia equivalía a positivismo, racionalismo, empirismo, materialismo mecanicista, tecnología, escepticismo, dogmatismo y especialización».

Mucho antes de que los nazis llegaran al gobierno algunos científicos alemanes y otros estudiosos exigieron que se creara una nueva ciencia que sustituyera a la existente, que ellos ponían en tela de juicio. Sería una nueva «ciencia aria», basada en conceptos intuitivos antes que en aquellos derivados de la teoría; en el éter, considerado residencia del *Geist* (espíritu); en el rechazo a aceptar concepciones abstractas o formalistas, vilipendiadas como señales de identidad de la «ciencia judía», y en la adopción, en la medida de lo posible, de avances científicos «obra de alemanes».

En un estudio ya clásico,¹⁹ Alan Beyerchen identifica algunos de los pilares básicos de la ciencia aria. Entre ellos encontramos temas inquietantemente similares a los que están en boga en la actualidad. Uno de los principios básicos de la ciencia aria era, por supuesto, que la ciencia, como dirían algunos, es en esencia un constructo social, de manera que la herencia racial del observador «afecta directamente la perspectiva de su obra». Científicos de razas

indeseables, por lo tanto, no estaban cualificados para trabajar. Más bien debían dedicarse a escuchar a aquellos que estaban en armonía con las masas, con el *Volk*. Además, este enfoque popular o *Völkisch* animaba a no expertos ideológicamente aleccionados a participar en juicios de valor sobre cuestiones técnicas (como los *Volksgesichte* o tribunales populares). El carácter internacional del mecanismo de consenso también resultaba abominable a los ideólogos nazis. El materialismo mecanicista, denunciado como fundamento del marxismo, debía ser erradicado de la práctica científica y la física debía reinterpretarse en relación no con la materia, sino con el espíritu. «Los defensores de la física aria descartaron así la objetividad y la internacionalidad en la ciencia [...] La objetividad en la ciencia no era más que un eslogan inventado por profesores universitarios para defender sus intereses». Hermann Rauschning, presidente del senado de Danzig, cita así unas palabras de Adolf Hitler:

Asistimos al fin de la Edad de la Razón [...] Una nueva era caracterizada por la explicación mágica del mundo está surgiendo, una explicación basada en la voluntad antes que en el conocimiento: la verdad no existe, ni en el sentido moral ni en el científico [...] La ciencia es un fenómeno social, y por tanto está limitada por la utilidad o por el daño que produce. Con su eslogan sobre la objetividad de la ciencia, la comunidad universitaria sólo busca liberarse de la tan necesaria supervisión del Estado.

Eso que llaman crisis de la ciencia no es más que los hombres empiezan a descubrir por sí mismos que han tomado el camino equivocado abrazando la objetividad y la autonomía. La sencilla pregunta que antecede a toda empresa científica es: ¿quién quiere saber algo, quién quiere orientarse en el mundo que le rodea?²⁰

También estaba la cuestión de cómo podía la tecnología, tal útil para el Estado, encajar en la idea romántica. En épocas recientes muchos movimientos antimodernistas, incluyendo algunos fundamentalistas, han abrazado la tecnología. Pero Philipp Lenar, un destacado físico a la par que héroe cultural de la propaganda nazi, hablaba en representación de al menos una minoría cuando afirmaba que la tendencia de los resultados científicos a preparar el terreno para avances prácticos ha conducido a una peligrosa noción, aquella del «dominio» del hombre sobre la naturaleza. Dicha actitud, sostenía, tan sólo ponía de manifiesto la influencia de «grandes técnicos espiritualmente empobrecidos» y de su «espíritu alienado y esencialmente destructivo». Esta idea hundía también sus raíces en la secular trayectoria de la formación del pensamiento romántico. Alan Beyerchen resume este hecho con la observación de que «el rechazo romántico al materialismo mecánico, al racionalismo, a la teoría y al pensamiento abstracto, a la objetividad y a la especialización llevaba tiempo relacionado con la creencia en un universo orgánico, con el énfasis en el misterio [y] en la subjetividad».

Puesto que todos estos excesos venían envueltos en un discurso en el que resuenan los ecos del empleado en la actualidad para deslegitimar la autoridad intelectual de la ciencia, es necesario recordar que hay un único ancestro común a estas visiones del mundo, antes que una relación

causal entre ellas. Ello es cierto también en el caso que presentaré a continuación, la postura adoptada por otro distinguido icono del humanismo contemporáneo, aunque tiene más de adalid que de analista real. Sus escritos sobre el tema son —como los de Ostwald Spengler o los positivistas— de interés aquí no porque representen posturas mayoritarias, que no es el caso, sino porque encierran el potencial de alcanzar una amplia resonancia llegado el caso de un punto de inflexión en el sentir de la época. También aquí veremos que la relación entre las ciencias de la naturaleza modernas y el ascenso de los totalitarismos, que Isaiah Berlin consideraba tan sólo el producto de una obscena reacción histórica, ahora es objeto de una interpretación mucho más siniestra: ambos fenómenos han pasado a vincularse de forma directa y causal.

Este inquietante vínculo ha sido apuntado repetidamente en los escritos del poeta, dramaturgo, combatiente en la resistencia contra la opresión marxista-leninista y estadista checo Václav Havel. En los pasajes que examinaremos aquí comprobaremos que Havel suscribe muchos de los puntos discutidos en el análisis de Isaiah Berlin; pero la idea central que propone es que el totalitarismo del siglo xx no fue más que el perverso resultado de una tendencia ideológica encarnada en el programa científico mismo. En este sentido, la ciencia occidental habría hecho posible el nacimiento del comunismo y, con la caída de éste, se ha visto gravemente comprometida.

Al repasar la historia del siglo xx, otros pensadores centroeuropeos podrían muy bien definirla como la liberación de las fuerzas de la brutalidad y la bestialidad irracionales, una reversión de las despiadadas autocracias en las que el destino de millones de individuos dependía de los caprichos del káiser Guillermo, de Hitler, Stalin y de sus secuaces, en lugar de ser el producto de un escepticismo organizado y de la búsqueda de consenso razonado, que están en el corazón de la ciencia. Pero para Havel las principales causas de los conflictos del siglo xx han de buscarse en precisamente lo opuesto: en el hábito —dicho en sus propias palabras— de «pensamiento racional y cognitivo», en la «objetividad deshumanizada» y en el «culto a la objetividad». Nos recomienda buscar refugio en la experiencia personal intransferible, en la intuición y el misterio y en los otros rasgos principales de la rebelión romántica. Debo dejarle defender su causa con sus propias palabras, porque si bien evita la documentación o el relato equilibrado propios del estudioso, domina muy bien el arte de la persuasión, de la suspensión de la incredulidad. El efecto en muchos de sus lectores es la aceptación hipnótica sin cuestionar las posibles generalizaciones o lagunas presentes del discurso. «El fin del comunismo», escribe Havel en uno de sus ensayos más citados:

... ha traído el final no sólo de los siglos xix y xx, sino de la era moderna en su conjunto. La era moderna ha estado dominada por la creciente creencia, expresada de distintas maneras, de que el mundo —y por tanto el Ser— es un sistema totalmen-

20

Hermann Rauschning. *Gespräche mit Hitler*. Nueva York: Europa Verlag, 1940, p. 210. Mussolini se expresaba de forma similar. [ed. esp. *Hitler, confesiones íntimas 1932-1934*, Barcelona: Circulo Latino, 2006].

te reconocible gobernado por un número finito de leyes universales que el hombre es capaz de comprender y moldear en beneficio propio. Esta era, que comenzó en el Renacimiento y se desarrolló desde la Ilustración hasta el socialismo, desde el positivismo al cientificismo, desde la Revolución Industrial hasta la revolución de la información, se caracterizó por el rápido avance del pensamiento racional, cognitivo. Ello a su vez desembocó en la orgullosa creencia de que el hombre, en tanto pináculo de todo lo que existe, era capaz de describir de forma objetiva, de explicar y controlar todo lo que existe, y de poseer la única verdad sobre el mundo. Fue ésta una era en la que se dio el culto a la objetividad despersonalizada, una era en la que el conocimiento objetivo se amasó y explotó tecnológicamente, una era de sistemas, instituciones, mecanismos y estadísticas. Fue una era de información gratuita, transferible, existencial. Una era de ideologías, doctrinas, interpretaciones de la realidad, una era en la que el objetivo último consistía en encontrar una teoría universal del mundo, y con ella, la llave a la prosperidad universal.

El comunismo fue el resultado de una perversión extrema de esta tendencia [...] La caída del comunismo puede interpretarse como una señal de que el pensamiento moderno —que se basa en la premisa de que el mundo es susceptible de ser objetivamente conocido— ha llegado a su crisis última. Esta era ha creado la primera civilización global o planetaria, técnica, pero ha llegado al límite de sus posibilidades, al punto a partir del cual empieza el abismo.

La ciencia tradicional, con su característica frialdad, puede describir las maneras diferentes en que podemos destruirnos a nosotros mismos, pero no es capaz de ofrecer instrucciones verdaderamente efectivas y practicables para evitar que lo hagamos.²¹

Llegado este punto alguien podría argumentar que las ideas aquí expresadas están construidas a partir de generalizaciones excesivas y saltos temporales ilógicos, que adolecen de los mismos defectos que las ideas de los monistas extremos; o al menos que la autodesignación de la ideología comunista como «científica» fue ciertamente un fraude. Sobre este último punto el estudioso de la historia y filosofía de la ciencia durante la Unión Soviética Loren Graham hizo la siguiente mordaz observación: «En 1992 el dramaturgo y presidente de la República Independiente de Checoslovaquia, Václav Havel, escribió que la caída del comunismo marcó el fin de una era, la desaparición del pensamiento basado en la objetividad científica [...] ¿Acaso la construcción del canal mar Blanco-Báltico en el lugar equivocado, con métodos primitivos y que causó la muerte a cientos de miles de prisioneros simboliza el auge de la racionalidad? ¿Acaso ignorar los consejos de los mejores técnicos del momento en la construcción de Magnitogorsk, el embalse de Dniéper y la línea férrea de Baikal-Amur supuso una victoria similar de la objetividad? ¿Fue la formación del mayor ejército de ingenieros jamás visto en el mundo, personas que llegarían a controlar por completo la burocracia soviética y que no sabían prácticamente nada de economía o de política un logro científico? [...] e incluso mucho tiempo después de muerto Stalin, entrada ya la década de 1980, ¿qué fue la insistencia soviética por mantener un sistema de granjas estatales ineficaces y gigantescas fábricas estatales sino una expresión de dogmatismo voluntarista que hacía caso omiso a una auténtica montaña de datos empíricos?».²²

Pero cabe dudar de si Havel reconsideraría su postura, ya que el objeto de su ensayo es, nada menos, presentar «el camino de salida de la crisis del objetivismo», como él la llama. Sólo un cambio radical en la actitud del hombre hacia el mundo servirá. En lugar de los métodos que, mediante la generalización y la objetivización, son capaces de producir explicaciones compartibles, repetibles, inter o trans subjetivas, debemos volver ahora la vista al extremo contrario, que, presumiblemente «la ciencia» ha desterrado por completo de este mundo; es decir, a «fuerzas como la experiencia natural, única e irreplicable del mundo, al sentido elemental de la justicia, a la capacidad de ponernos en el lugar de los demás [...] Coraje, compasión y fe en la importancia de medidas particulares que no aspiran a convertirse en la llave a la salvación universal [...] Debemos ver el pluralismo presente en el mundo [...] Debemos esforzarnos más para comprender que para explicar». El hombre necesita «espiritualidad individual, conocimiento personal y de primera mano de las cosas [...] y, sobre todo, confiar en su propia subjetividad como su lazo principal de unión con la subjetividad del mundo...».

A pesar de que Havel alude, como de pasada, a una posible fusión entre «la construcción de soluciones sistémicas universales» o a «la representación y el análisis científico» con la autoridad de la «experiencia personal» para así conseguir dotar a la política de «un rostro nuevo y posmoderno», su anuncio del «Fin de la era moderna» no debe ser entendido únicamente como una llamada al compromiso o a la coexistencia entre planteamientos rivales; así lo aclaraba en una versión anterior y más afilada de este ensayo, en la que trataba sobre el lugar de la ciencia moderna sin ninguna clase de ambigüedades y que por tanto merece leerse con atención:

[La nuestra] es una época que niega la importancia determinante de la experiencia personal —incluida la experiencia del misterio y de lo absoluto— y que sustituye la experiencia del misterio y de lo absoluto con una nueva visión del mundo, hecha exclusivamente por el hombre, libre de misterio, libre de los «caprichos» de la subjetividad y, por tanto, impersonal y deshumanizada. Es el absoluto de la llamada objetividad, la cognición racional del modelo científico del mundo.

La ciencia moderna, en su construcción de una imagen del mundo universalmente válida, rompe así los límites del mundo natural, el cual concibe sólo como una cárcel de prejuicios de los cuales debemos desprendernos para encontrar la luz de la verdad comprobada científicamente [...] Con ello por supuesto está destruyendo, puesto que los considera una ficción, los cimientos más íntimos de nuestro mundo natural. Mata a nuestro Dios y ocupa su trono vacante, de manera que a partir de ese momento será la ciencia la que controlará el orden de la naturaleza, en tanto que su único legítimo guardián y en tanto también que único legítimo árbitro de la verdad relevante. Porque, después de todo, sólo la ciencia es capaz de alzarse por encima de las verdades individuales subjetivas y sustituirlas con una verdad superior, transubjetiva y transpersonal que sea auténticamente objetiva y universal.

El racionalismo y las ciencias modernas, a través del trabajo del hombre que, como en todas las tareas humanas, se desarrolla dentro del mundo natural, pueden ahora dejarlo [el mundo natural] atrás, degradarlo y difamarlo, y, por supuesto, al mismo tiempo colonizarlo.²³

21

«Politics and the World Itself». *Kettering Review* (verano de 1992), pp. 9-11. Este ensayo también se publicó el 1 de marzo 1992 en el *New York Times* a modo de editorial y bajo el título de «The End of the Modern Era».

22

Loren R. Graham. *The Ghost of the Executed Engineer: Technology and the Fall of the Soviet Union*. Cambridge, Massachusetts: Harvard University Press, 1993.

23

Reeditado en Jan Vladislav (ed.). *Vaclav Havel, or Living in the Truth*, Faber and Faber, Londres, 1987, pp. 138-139. Este párrafo es de 1984.

24

El 4 de julio de 1994 Havel aprovechó la ocasión para reiterar en gran medida su argumento anterior, con la excusa de explicar la actual «mentalidad que llaman posmoderna» y «la crisis» a la que la ciencia ha llevado a la humanidad. La única novedad de este discurso (publicado en forma de editorial el 8 de julio en el *New York Times*) es que nuestra «integridad perdida» puede ser renovada, paradójicamente, por una «ciencia que sea nueva, posmoderna», ejemplos de la cual serían «el principio cosmológico antrópico» o «la hipótesis de Gaia». Esto resultó demasiado incluso para Nicholas Wade, quien escribió un ataque devastador al ensayo de Havel (N. Wade. «Method and Madness: A Fable for Fleas», en el *New York Times* el 14 de agosto de 1994, p. 18), que terminaba diciendo: «Una concepción del mundo construida sobre el principio antrópico y la hipótesis de Gaia no sería ciencia posmoderna sino más bien una vuelta atrás a la numerología y astrología, de las cuales el racionalismo aún no nos ha rescatado por completo [...] Subvertir el significado del racionalismo transformándolo en misticismo sería una cura aún más perniciosa que la enfermedad». La tentación de ser incluidos en las filas de la posmodernidad parece haber atraído incluso a científicos; el mejor ejemplo de ello es el interés posmoderno de éstos en los llamados «límites de la ciencia». El encendido debate sobre este tema se inició nada menos que en la década de 1870 con Emile Dubois-Reymond y también preocupó a los positivistas lógicos. Para otros ejemplos de este ya viejo problema véase G. Holton y R. S. Morison (eds.). *Limits of Scientific Inquiry*. Nueva York: W. W. Norton, 1978.

25

Publicado en septiembre de 1992 en *American Journal of Physics*, vol. 60, núm. 9: 779-781.

26

Hay una grabación de la sesión (12 de febrero de 1993) disponible en la American Association for the Advancement of Science. Los «Comentarios introductorios», realizados por el mismo George Brown, también se distribuyeron con una nota de prensa de su oficina en Washington.

27

En la reunión anual del 12 de febrero de 1993 de la American Association for the Advancement of Science.

Vemos aquí el paso de gigante con el que Havel ha dejado atrás el análisis de Berlin: lo que ha matado la edad moderna es la ciencia moderna en sí misma. Como en respuesta a los excesos de Ostwald, se la considera responsable incluso de deicidio.

Muchos se han sentido conmovidos por la poderosa combinación de Havel de sentimiento poético, florituras teatrales y la manera en que blande con audacia su vieja camisa ensangrentada de víctima de la persecución política. El resumen de sus ideas, publicado con el conspicuo título de «El fin de la era moderna»²⁴ tuvo una aceptación inmediata y acritica entre lectores de ideas diversas. Entre ellos estaba una persona particularmente bien situada para ponderar los valores de la ciencia y para sacar conclusiones de importancia para la vida científica de Estados Unidos. Estamos recorriendo ya las últimas etapas en nuestro camino hacia la comprensión del lugar actual de la ciencia en nuestra cultura.

Aquella persona en quien el escrito de Havel dejó tan honda impresión no era otro que el distinguido presidente del comité del Congreso de Estados Unidos para Ciencia, Espacio y Tecnología y uno de los más aguerridos y también eficaces defensores de la ciencia durante su prolongada estancia en la Casa de Representantes: George E. Brown, Jr, congresista por el estado de California, quien, tras reconocer haberse «sentido inspirado» por el ensayo de Havel «El fin de la era moderna», decidió reconsiderar su papel como público defensor de la ciencia. En primer lugar escribió un ensayo largo e introspectivo²⁵ titulado «La crisis de la objetividad» y a continuación lo presentó a un grupo de científicos sociales en una sesión pública en la reunión anual de la Asociación Estadounidense para el Avance de la Ciencia (American Association for the Advancement of Science), con el título de «La crisis de la objetividad. Repensar el papel de la ciencia en la sociedad».²⁶

Persuadido por la versión de Havel de la rebelión romántica, Brown se dedicó a estudiar con afán las consecuencias que ésta debería tener en la práctica científica en Estados Unidos. En su calidad de pragmático líder político, le preocupaba sobre todo cómo podía preservarse la legitimidad de la actividad científica, sirviendo a la nación en términos de «avances sostenibles en la calidad de vida», «el deseo de justicia» (que, afirma, «es considerado como algo externo al ámbito del análisis científico») y el resto de los problemas «reales y subjetivos que afectan a la humanidad». Para entonces veía escasos indicios de que «el conocimiento científico objetivo tenga beneficios subjetivos para la humanidad». La reivindicación de libertad para llevar a cabo investigación básica sin restricciones políticas y morales es inútil también, dijo, porque toda investigación es «contextual» y está sujeta «al curso de la historia».

Además, la ciencia ha usurpado la primacía «sobre otras clases de cognición y experiencia». Aquí Brown citaba la definición que hacía Havel de la «crisis de la objetividad» como resultado de la supuesta subyugación de nuestra

humanidad subjetiva, de nuestro «sentido de la justicia [...] sabiduría arquetípica, buen gusto, coraje, compasión y fe». Los procesos de la ciencia «no sólo no nos pueden ayudar a distinguir entre el bien y el mal, sino que además –afirma– sus resultados son ciertos o no independientemente de su valor moral». En suma, sostenía Brown, sería demasiado fácil apoyar la investigación científica sin más cuando la solución es «cambiarnos a nosotros mismos». Ciertamente llegó a la conclusión de que «la promesa de la ciencia puede muy bien estar en la raíz de nuestros problemas». Claro está que los científicos continuarán siendo de utilidad, aunque sus investigaciones habrán de ser convenientemente dirigidas al campo de la educación o a trabajar con vistas «a metas específicas que determinen un contexto general para la investigación», tales como el control demográfico. Brown abrazaba así una suerte de baconismo y rechazaba la visión más general de la ciencia de Vannevar Bush, un rechazo al que ya he hecho referencia hacia el comienzo de este ensayo (véase nota 2). Al igual que Havel, la respuesta de Brown a la cuestión de si la ciencia puede ocupar un lugar central en la cultura moderna era, claramente que *no*.

Cuando Gordon Brown presentó sus ideas a un público de expertos en la sesión que había organizado y para la cual había seleccionado un plantel de científicos sociales,²⁷ tan sólo uno de éstos se atrevió a mostrar abiertamente su desacuerdo, mientras que otro incluso urgió a Brown a ir más allá: tal vez sin darse cuenta de hasta qué punto se estaba acercando a la solución *völkische* ensayada anteriormente en otros momentos y lugares, incluida la revolución cultural de Mao, sugirió seriamente que, con objeto de filtrar las subvenciones federales a los proyectos de investigación científica, se formara una variante de la National Science Foundation Board (Comité Nacional para la Ciencia), cuyos miembros deberían incluir a no expertos tales como «un sin techo y el miembro de una banda urbana». Ninguno de los asistentes hizo la más mínima objeción audible. Era como tener una visión del futuro. Pero también es importante señalar que más tarde el señor Brown, aparentemente movido por las objeciones intelectuales a sus teorías tales como las expresadas anteriormente y que le transmitieron uno o dos científicos, se distanció de la postura de Havel. Ciertamente uno no puede menos que coincidir con él en que en el contexto inmediatamente posterior a la Guerra Fría resulta «más imperativo alistar a la ciencia y a la tecnología en una campaña en pos de una sociedad más humana y productiva, en la que todos los estadounidenses puedan disfrutar de los beneficios de una calidad de vida mejor».²⁸

En este breve recorrido, desde los trémulos pilares de la tradición platónica en Occidente hasta el llamado «fin de la era moderna» y «fin del progreso», hemos identificado algunas de las tendencias históricas principales que han surgido y caído y resurgido de nuevo dentro de la amalgama de ideas de la cual emerge la visión dominante de una época. La versión actual de la rebelión romántica, aunque

28

George E. Brown, «New Ways of Looking at U.S. Science and Technology». *Physics Today*, 47: 32, 1994. En el transcurso de una charla sobre «The Roles and Responsibilities of Science in Post-Modern Culture» (El papel y la responsabilidad de la ciencia en la cultura posmoderna, 20 de febrero de 1994, con motivo de otra reunión anual de la American Association for the Advancement of Science), el señor Brown dijo: «Déjenme empezar sugiriendo que el término "cultura posmoderna", en ocasiones empleado para describir la era actual, es una rúbrica que procede de las artes y la arquitectura, donde tiene un significado concreto. En mi opinión, utilizar el término posmoderno para definir un período en política, legislación o economía sólo conduce a confusión, y no nos ayudaría a definir un punto de partida para nuestro debate de hoy. Espero que el discurso hoy no derive en una tediosa disección del posmodernismo. Debo señalar, no obstante, que el editorial publicado en el *New York Times* hace dos años titulado "El fin de la era moderna" por el filósofo y dramaturgo checo Václav Havel contenía varios puntos con los que estoy de acuerdo y que he incluido en charlas mías anteriores. Aunque Havel llega hasta los términos modernismo y posmodernismo por vía de su formación y experiencia artísticas, yo no suscribo esas etiquetas, en gran medida porque no comprendo del todo el uso que hace de ellas». El señor Brown es también uno de los pocos legisladores que ha protestado por el reciente edicto de la senadora Barbara Mikulski solicitando que la subvención de proyectos de investigación básica «inspirada sólo en la curiosidad» sea reducida a favor de una supuesta «investigación estratégica».

29

Véase sobre todo Harvey Brooks, «Research Universities and the Social Contract for Science», en Lewis Branscomb (ed.), *Empowering Technology: Implementing a U.S. Strategy*. Cambridge Massachusetts: MIT Press 1993. Brooks ha sido uno de los autores que mejor han observado y predicho el lugar de la ciencia en nuestra cultura. Véase por ejemplo su ensayo «Can Science Survive in the Modern Age?», en *Science*, 174: 21-30, 1971.

30

Véase por ejemplo Don. K. Price, «Purists and Politicians», *Science*, 163: 25-31, 3 de enero de 1969.

poderosa en otros campos, representa sólo una seductora pero minoritaria visión dentro de los analistas y legisladores; no nace de las raíces sino de las copas de los árboles. Sin embargo, aunque la sostienen personas de relevancia pública que pueden ciertamente influir en la orientación de un cambio cultural, la reacción de los científicos en general y de las altas esferas científicas en particular ha sido de silenciosa aquiescencia. Si estas tendencias continúan y aquellos que se llaman a sí mismos posmodernos pasan a ser la fuerza dominante, la nueva sensibilidad de la era que empieza será ciertamente muy distinta de la que ha dominado hasta ahora.

Los expertos en ciencia política debaten en la actualidad lo que ellos llaman la renegociación en curso del «contrato social» entre ciencia y sociedad.²⁹ Se podría argumentar que este cambio se ha retrasado por muchas razones, una de ellas que el estatus relativamente protegido de que ha gozado la ciencia durante muchas décadas no tenía tanto que ver con el compromiso de la sociedad como con la Guerra Fría y la amenaza de sus posibles repercusiones, que, como advirtió ya hace tiempo Don K. Price,³⁰ con el tiempo se volverían en contra de lo científicos. Si a ello sumamos preocupaciones por el estado de la economía y la competitividad, la ausencia de una educación científica de calidad, etcétera, no resulta difícil explicar la disposición del público a reconsiderar el papel de la ciencia en la sociedad. Pero en mi opinión estos factores son meros catalizadores del cambio de opi-

nión generalizado que siempre está potencialmente presente en nuestra cultura.

Por supuesto es posible que esta versión reciente de la rebelión romántica se agote por sí sola, aunque dudo que sea así. También puede ganar fuerza, como ocurrió en el siglo XIX y de nuevo en distintos momentos del XX, en especial cuando la atención de la comunidad científica se apartó del curso de los acontecimientos. O quizá, con un poco de suerte surja una nueva tendencia, una «tercera vía» basada en un concepto análogo a la complementariedad (y también análogo a la complementariedad entre ciencia personal y pública dentro de la práctica misma de la investigación). Es decir, que al menos se reconozca, tanto por parte de los intelectuales como del público general, que los aspectos científicos y humanísticos de nuestra cultura no tienen por qué representar visiones del mundo contrapuestas y mutuamente excluyentes, sino que son elementos complementarios que pueden coexistir y de hecho coexisten de manera productiva (o, como tan bien resumió Samuel Taylor Coleridge en el capítulo XIV de su *Biografía literaria*: «En el equilibrio o reconciliación de cualidades opuestas o discordantes»). En todo caso, los historiadores asistirán a las próximas fases en la ya antigua batalla por definir el lugar de la ciencia en nuestra cultura con creciente fascinación, aunque quizá también con el incómodo recuerdo de la profecía de Ostwald Spengler, o el pesimismo de Sigmund Freud, o el sombrío análisis que hacía Isaiah Berlin de nuestra era moderna.

estructura y evolución del universo

FRANCISCO SÁNCHEZ MARTÍNEZ

La gran aventura

Desde que el hombre es *sapiens* ha mirado e interrogando al cielo con sumo interés y sobrecogimiento. Cosa bien natural puesto que lo que sucede sobre nuestras cabezas nos afecta y mucho. Los hombres primitivos palpaban de manera directa la influencia de los cielos en sus vidas, tan expuestos como estaban a la naturaleza, y tan dependientes de la muerte y resurrección del Sol, de la Luna y las estrellas que marcaban los ritmos de días y noches, y de las estaciones. Ellos aprendieron a predecir mirando la situación de los astros, pues necesitaban conocer lo que iba a pasar para poder procurarse la comida y evitar, a su vez, ser comidos por los otros predadores. Imaginemos, por otra parte, el estupor y el espanto que producirían a nuestros ancestros los imprevistos y dramáticos fenómenos que observaban en las alturas: rayos, truenos, auroras polares, estrellas fugaces, bólidos, cometas, eclipses de Sol y de Luna. ¿Cómo no ver en ellos señales de seres muy superiores? Lógico que pusiesen en los cielos la morada de sus dioses. Algunos se dieron cuenta enseguida de cómo el conocer los secretos de tales fenómenos, haciendo creer a los demás en su capacidad de usarlos en beneficio o maleficio de unos y otros, les confería gran poder y carácter de mediadores divinos. Por eso, aun en las civilizaciones más arcaicas, aparecen mitos, ritos y augurios celestes, guardados por las castas sacerdotales. Pero aún hoy, en las más desarrolladas y tecnificadas sociedades de nuestro siglo XXI, colean estos primitivismos en forma de astrolo-

gía, sectas astrales y demás supercherías. Y hasta las teorías y modelos cosmológicos científicos más complejos y especulativos, son vividos y defendidos por muchas de las personas cultas de nuestro mundo con fanatismo religioso. Todo esto hay que tenerlo presente cuando se trate de hacer, como yo ahora, una puesta al día, necesariamente sintética y asequible, de lo que en estos momentos sabemos sobre la estructura y evolución del desmesurado Universo al que pertenecemos.

Lo primero que hay que recordar y subrayar, para poner en su sitio lo que diré a continuación, es que todo conocimiento científico es provisional y está permanentemente sujeto a revisión. Y que, además, lo que voy a contar son especulaciones con base científica rigurosa, pero especulaciones al fin, que tratan de dar explicación a lo que hasta estos momentos hemos podido observar con los telescopios e instrumentos más avanzados. La realidad completa del inconmensurable Cosmos nos es inasequible, por nuestra propia finitud. En definitiva, que siendo mucho, muchísimo, lo que sabemos, es muchísimo más lo que aún ignoramos. He tenido que vencer la tentación de escribir en esta puesta al día, junto a cada aseveración, el racimo de preguntas incontestadas que la envuelven. Estas profundizaciones hay que dejarlas para libros especializados.

Dicho todo esto, aún resulta más emocionante contemplar la aventura inacabada y bella de los humanos avanzando a ciegas, pero decididos, tras los misterios infinitos, impelidos por su curiosidad y su afán innato por conocer

Página anterior:

Figura 1. En esta imagen dos galaxias espirales –NGC2207 e IC2163– se encuentran en su fase inicial de interacción. Con el tiempo, ambas galaxias terminarán formando una sola. Las fuerzas de marea gravitatoria de NGC2207 han torcido la forma de IC 163, que expulsa estrellas y vierte gases en serpentinadas largas que sobresalen cien mil años-luz por el borde derecho de la imagen. Así se han ido formando las grandes galaxias espirales. Nuestra propia galaxia, la Vía Láctea, ha asimilado en el pasado galaxias menores y está engullendo actualmente otras. En un futuro llegaremos a chocar con M31, la mayor de las galaxias espirales de nuestro Grupo Local. (Créditos: NASA)

que nos ha hecho capaces de superar nuestras fuertes limitaciones hasta cotas impensables. Por ejemplo, en el caso concreto del sentido de la vista, siendo unos bichos bastante cegatos, incapaces de distinguir objetos lejanos, y limitadísimos en el rango de longitudes de onda del espectro electromagnético que nos es asequible, hemos sido capaces de inventar unas «prótesis» maravillosas, que son los telescopios, para lograr «ver» objetos celestes que están a distancias tan enormes que tenemos que medirlas en miles de millones de años luz.

Según ha ido aumentando nuestra capacidad de «ver», de observar más lejos y con más detalle, nuestra idea del Cosmos ha ido cambiando, y lo ha hecho drásticamente en los últimos tiempos (figura 2). Por no remontarnos demasiado en la historia, retrocedamos sólo hasta el Renacimiento, a la época de la «revolución copernicana», que quitó a la Tierra del centro del Universo, después de reducirla a un cuerpo redondo girando en torno al Sol, y nos mostró que el material celeste era de nuestra misma naturaleza. No más divino que el propio polvo terrestre. Durante unos pocos siglos anduvimos maravillados con estas cosas tan insólitas y fantásticas, y entretenidos con la precisa mecánica celeste estudiando los movimientos de los cuerpos de nuestro Sistema Solar. Consecuentemente, se creía en la perfección del reloj cósmico y la inmutabilidad del Universo.

Solo hace menos de un siglo, toda esta perfección y serenidad se nos vino abajo. Confirmada científicamente la existencia de otros muchos «universos islas» situados fuera de nuestra Vía Láctea —hasta entonces el único enjambre de estrellas y nebulosas rodeado por el vacío infinito—, y que la separación entre ellos aumentaba más velozmente cuanto más lejos estaban, se tuvo que aceptar la expansión del Universo: un Universo que aumenta de tamaño y se va enfriando. Bastaba dar marcha atrás a tal expansión para llegar a la singularidad original, y de ahí a la Gran Explosión origen de todo. Añadiendo a esto el descubrimiento de la energía que hace brillar las estrellas, se tuvo que aceptar que el Universo entero está en permanente evolución. Ni estático ni eterno, en él todo está «vivo», también las estrellas y las galaxias. Por eso vemos por doquier nacer y morir objetos celestes, en procesos de transformación y reciclaje perpetuos.

Y de repente, nos llega la sorpresa de un Universo en expansión ¡acelerada! Y de nuevo patas arriba la cosmología. Este aceleramiento global fuerza a recomponer de arriba abajo el edificio de la física cósmica, y a pensar en algún tipo de energía misteriosa, ligada a lo que llamamos vacío, que lo llena todo y lo empuja todo. Energía que produce fuerzas antigravitatorias capaces de contrarrestar la implosión prevista y, además, con enorme poder para acelerar la expansión del espacio.

Vamos ahora juntos a recorrer la sendita abierta por la ciencia en el conocimiento de la estructura y evolución del Universo al que pertenecemos.

El Universo se acelera

Uno de los grandes problemas prácticos para explorar los confines del Universo es la determinación precisa de distancias. Sin entrar en tecnicismos, diremos que con las supernovas tipo Ia —estrellas terminales cuyo brillo intrínseco es prácticamente igual para todas ellas— usándolas como «candelas» patrones se consigue poner distancia a galaxias lejanísimas. El concepto es simple: si todas brillan igual en origen, el brillo aparente que observamos desde la Tierra nos da la distancia a la que se encuentra la supernova y, naturalmente, la de la galaxia a la que pertenece.

Por este camino, los grupos de investigación dirigidos por Saul Perlmutter y Adan Riess, pudieron determinar, independientemente, las distancias de las galaxias donde se producían estos tipos de explosiones de supernova, y comparando los datos así obtenidos con los corrimientos al rojo de dichas galaxias —que miden la expansión del Universo—, se toparon con la sorpresa de que dicha expansión era más lenta cuanto más lejanas estaban las galaxias. O dicho de otra forma: el Universo se expandía más despacio en el pasado que hoy día. En fin, que el Universo está gobernado actualmente por una expansión acelerada.

Estas observaciones de los grupos de Perlmutter y Riess constituyen el primer dato observacional de que el ritmo de expansión del Universo no ha sido el mismo a lo largo de su prolongada historia. Hecho de enorme trascendencia, como luego veremos.

Pero antes de seguir adelante, detengámonos un momento en el concepto de expansión del Universo, el cual, aun siendo uno de los conceptos fundamentales de la ciencia moderna, es uno de los peor entendidos. La mayoría de la gente se imagina que la Gran Explosión es como la de una bomba gigantesca, que explosiona en un punto del espacio y, desde él, la materia resulta expelida hacia afuera, debido a las diferencias de presión en la deflagración. Pero la Gran Explosión de que hablamos los astrofísicos no fue una explosión «en el espacio», fue una explosión «del espacio», matiz este verdaderamente importante. En este tipo de explosión tan peculiar, densidad y presión se mantienen constantes en el espacio, aunque decrecientes en el tiempo.

Se suele emplear para visualizar esto de la explosión del espacio el símil del globo que, al inflarse, hace que los detalles de cualquier cosa que esté sobre su superficie se vean separándose unos de otros: en la superficie del globo, todo se aleja de todo. Esta imagen bidimensional es muy gráfica, pero tiene el problema de que puede hacer creer que, al igual que en el globo, también hay un centro, un punto, de donde dimana la Gran Explosión. Pero resulta que la Gran Explosión sucedió en todos los puntos del espacio a la vez, no en ninguno determinado, de acuerdo con la teoría general de la relatividad de Einstein. Según los modelos más aceptados, el Universo no requiere ni un centro desde el que dilatarse, ni espacio vacío que lo envuelva para expandirse. Ni siquiera necesita de más de



Figura 2. Así vemos el Universo lejano mediante una «lente gravitatoria». El cúmulo de galaxias Abell 2218 es tan masivo que es capaz de curvar los rayos de luz creando imágenes a la manera de una lente. Los arcos que apreciamos son imágenes distorsionadas de galaxias mucho más lejanas que el propio cúmulo. Este efecto amplificador permite penetrar profundamente en el Cosmos, ver más lejos. (Créditos: NASA)

tres dimensiones, pese a que algunas teorías, como la teoría de cuerdas, precisan algunas más. Su base teórica es la Relatividad, la cual establece que al espacio le bastan tres dimensiones para poder expandirse, contraerse y curvarse. Por otra parte, no debe imaginarse que la singularidad que dio origen a la Gran Explosión fue algo pequeño, el «átomo inicial» del que se habla a veces. Y ésta se genera independientemente de cual sea el tamaño del Universo, ya sea éste finito o infinito.

Recordemos ahora que los átomos emiten y absorben luz a ciertas longitudes de onda bien concretas, estén en un laboratorio o en las galaxias más lejanas. Pero en este segundo caso las observamos desplazadas hacia longitudes de onda mayores —corrimiento hacia el rojo—. Lo que sucede es que, a medida que el espacio se expande, las ondas electromagnéticas se estiran, o sea: se hacen más rojas. Y este efecto permite medir la velocidad con que se separan las galaxias unas de otras, conocida como velocidad de recesión. Conviene subrayar que este desplazamiento al rojo cosmológico no es el corrimiento Doppler normal que se produce dentro del espacio, y sus fórmulas también son diferentes.

Pese a la expansión generalizada del espacio hay galaxias, como nuestra vecina Andrómeda, que se nos están

acercando, que parecen no seguir esta ley. Son excepciones aparentes, causadas porque en las cercanías de las grandes acumulaciones de materia prepondera la energía gravitatoria, que hace girar a estos gigantescos enjambres de estrellas unos en torno a otros. Las galaxias lejanas también presentan estos efectos dinámicos locales, pero desde nuestra perspectiva tan lejana, quedan encubiertos por sus grandes velocidades de recesión.

Para mayor complicación resulta que el Universo no sólo está expandiéndose, sino que además, lo hace aceleradamente. Situación que ha hecho resucitar la constante que Einstein introdujo en su Relatividad General para mantener el paradigma de Universo estacionario. Mientras se creía que vivíamos en un Universo en deceleración, era lógico pensar que con el tiempo podríamos observar más y más galaxias, pero en un Universo acelerado tiene que acontecer lo contrario. El horizonte cósmico de sucesos, determinado por la finitud de la velocidad de la luz y el aumento de la velocidad de recesión del espacio, marca una frontera, más allá de la cual sucederán fenómenos que nunca veremos, porque no nos puede alcanzar la información que emitan. Al ir acelerándose la expansión del espacio, paulatinamente irán desapareciendo de nuestra visión una galaxia tras otra, empezando por las más lejanas.

Energía oscura

El hecho del Universo acelerado ha cogido desprevenidos a astrónomos y físicos, quienes andan en plena efervescencia de especulaciones científicas para explicarlo. La imaginación humana es capaz de idear muchas teorías, y bien ingeniosas, pero sólo perdurarán aquellas que expliquen todas las observaciones. La observación astronómica sigue siendo la «piedra de toque» de cualquier teoría, y por elegante que ésta sea habrá de ser validada por las observaciones. Por ello, los científicos parecen lentos y conservadores, moviéndose con pies de plomo a la hora de cambiar las teorías consolidadas y los paradigmas.

Una forma inmediata de interpretar la expansión acelerada sería considerar que la gravedad no sigue la misma ley en nuestro entorno cercano que a escalas supergalácticas, y que a tales distancias no puede frenar la expansión puesto que su poder atractivo no se prolonga indefinidamente en la distancia. Otra propuesta, ya formulada, sería que la aceleración observada es una apariencia causada por el tiempo mismo que va gradualmente enlenteciéndose. Pero los cosmólogos prefieren mantener la universalidad de las leyes físicas establecidas en el planeta Tierra y sus alrededores, y se han lanzado a postular la existencia de una especie de fluido cósmico de propiedades contradictorias que lo llena todo y que se manifiesta como una energía desconocida —bautizada como energía oscura— que en vez de atraer repele, tan poderosa que vence muy eficazmente a la atracción gravitatoria de las enormes masas de los cúmulos y supercúmulos de galaxias.



Figura 3. El Very Small Array –IAC y las universidades de Cambridge y Manchester– es uno de los instrumentos instalados en el Observatorio del Teide del Instituto de Astrofísica de Canarias para medir anisotropías en el Fondo Cósmico de Microondas. En este observatorio se mide sistemáticamente desde hace más de veinte años esta radiación primigenia con diferentes instrumentos y técnicas. (Créditos: IAC)

Por el momento, las observaciones disponibles están a favor de esta energía oscura. Las primeras fueron, como anteriormente señalamos, las hechas mediante fotometría de las supernovas tipo Ia, y mostraron que las galaxias más viejas se están expandiendo a un ritmo más lento que el actual. Pero también las medidas de la radiación del Fondo Cósmico de Microondas —o radiación de fondo— apuntan en esta dirección.

Recordemos que fueron Penzias y Wilson quienes, en 1965, descubrieron esta radiación, este ruido de fondo que lo llena todo, y su descubrimiento sirvió para afianzar los modelos de Gran Explosión. Muy posteriormente se lograron detectar anisotropías en el Fondo Cósmico de Microondas, que aun siendo pequeñísimas —del orden de 0,00001%—, están llenas de información sobre los orígenes de la estructura del edificio gigantesco del Cosmos que vemos. Pues bien, parece requerirse la contribución de la energía oscura para completar la densidad del Universo obtenida a través de las medidas del Fondo Cósmico de Microondas. Resulta que, puesto que los tamaños de las irregularidades en la radiación de fondo son un reflejo de la geometría global del espacio, sirven para cuantificar la densidad del Universo, y esta densidad es netamente mayor que la simple suma de la materia ordinaria y la exótica. Pero además, las modificaciones causadas en esta radiación por los campos gravitatorios de las grandes estructuras cósmicas dependen de cómo ha cambiado el ritmo de expansión. Y este ritmo concuerda con las predicciones de los modelos de energía oscura.

La distribución de los enjambres de galaxias muestran ciertos patrones, concordantes con las «manchas» observadas en la radiación de fondo, que sirven para estimar la masa total del Universo (figura 3). Pues también estos modelos requieren de la energía oscura. Y los estudios de la distribución de las lentes gravitatorias —los objetos muy masivos, recordemos, se comportan como lentes curvando las trayectorias luminosas— parecen necesitar de la energía oscura para explicar el crecimiento en el tiempo de las aglomeraciones de materia. Pero no todo son confirmaciones, hay observaciones que no pueden explicar estos modelos, como por ejemplo las abundancias de los cúmulos de galaxias más lejanos.

Son muchos los investigadores que están tratando de encontrar en la energía oscura la causa de las cosas no explicadas en sus modelos previos. Los datos se van acumulando y los modelos son cada vez más refinados. No hay duda de que se irán añadiendo más indicios y hasta pruebas, tal es la fiebre causada en los científicos por esta sorprendente aceleración del Universo.

La omnipresencia de la energía oscura es tan sutil que, aun llenándolo todo nos ha pasado desapercibida, porque está muy diluida y porque no se acumula formando «grumos», como la materia. Para que sus efectos sean apreciables se necesitan espacios y tiempos muy grandes, pese a ser la energía más poderosa del Cosmos. Permítanme que

añada que en esta energía, que actúa como una fuerza repulsiva, lo que la hace tener presión negativa, caben dos posibilidades: la llamada «energía fantasma» y la llamada «quintaesencia». Todo ello muy sugerente, pero difícil de digerir científicamente hablando, ya que se trata de fuerzas que no entendemos y no podemos observar.

Resulta absolutamente lógico pensar, que si tal energía representa más de las tres cuartas partes de nuestro Universo, ha tenido que influir y mucho en todo su proceso evolutivo, determinando su estructura a gran escala y la formación de los cúmulos de galaxias. Hasta la evolución de las propias galaxias debe estar marcada por su omnipresencia. La formación de las galaxias y su agrupación en cúmulos sabemos que está determinada por las interacciones, choques y fusiones entre ellas —nuestra propia Vía Láctea se piensa que es el resultado de la coalescencia de quizá un millón de galaxias enanas— por lo que la energía oscura ha tenido que jugar un papel nada despreciable en todo ello. No obstante, la confirmación clara vendrá cuando seamos capaces de determinar si el comienzo del predominio de la expansión acelerada coincide en el tiempo con el final de la formación de las grandes galaxias y de los supercúmulos.

Universo en cuatro dimensiones

Le he dado muchas vueltas a cómo mostrar lo que sabemos hoy sobre la estructura del Universo al que pertenecemos. No resulta nada fácil por múltiples razones, y no tan sólo porque no sea sencillo sintetizar para los no especialistas, sin dejar cabos sueltos importantes que solemos dar por sobrentendidos.

Si consideramos como Universo todo lo que existe, desde las entidades mínimas a las más gigantescas, una forma de mostrar su estructura sería hacer un inventario de todos estos elementos y situarlos jerárquicamente en el espacio. Pero quedaría incompleto sin decir cuáles son sus interconexiones e interrelaciones. Además, todo ello —elementos e interconexiones— no son algo estático, y todo el conjunto está interaccionando y modificándose permanentemente. Démonos cuenta de que, asimismo, no podemos tener una «instantánea» de lo que hay en el Universo en este momento, pues cuando observamos con un telescopio en una dirección, nuestra mirada, al ir penetrando más y más profundamente, va retrocediendo más y más en el tiempo: obtenemos una cuña de la historia del Universo, no una instantánea. Sin embargo, puesto que todas las direcciones del Universo resultan estadísticamente idénticas, lo que vemos en cualquier dirección a una distancia de miles de millones de años luz debe ser una representación de nuestra propia región del espacio, y de cualquier otra, tal y como fueron hace miles de millones de años.

Iremos por etapas. Primero recordaremos que, de acuerdo con cuanto venimos diciendo, más de las tres cuartas partes de nuestro Cosmos, ahora, es una forma de energía misteriosa que llamamos energía oscura, y del resto, más

del 85% es materia oscura, que no podemos ver porque aunque interacciona gravitatoriamente, no lo hace con la radiación. O sea, que no mucho más del 3% de todo el Universo es materia ordinaria, de la que sólo acertamos a visualizar una pequeña parte de ella concentrada en estrellas y galaxias. Recordemos que llamamos materia ordinaria a la materia bariónica —protones, neutrones, etc.— de la que nosotros estamos hechos. Y esta materia la encontramos, en su mayor parte, en forma de plasma gaseoso ionizado, estando en forma sólida y líquida sólo una muy mínima parte de la misma. ¡Qué difícil es pensar que los inmensos océanos y el suelo firme de la corteza terrestre, que tanta seguridad nos da al pisarlo, son algo verdaderamente rarísimo en el Universo! Pero la ciencia nos ha llevado a la aceptación de que vivimos en un lugar bien exótico, situado en un rincón cualquiera del Cosmos.

El panorama, por otra parte, no puede dejar de ser más descorazonador: ¡no tenemos ni la más remota idea, pese a nuestras elegantes especulaciones científicas, de la naturaleza de un 97% de los constituyentes de nuestro Universo! Claro que saber esto, ya es un gran triunfo de la gran aventura humana en pos del conocer.

Por nuestra propia naturaleza, nos movemos y entendemos en tres dimensiones espaciales, más el tiempo. Y en este espacio-tiempo están desarrollados los modelos relativistas al uso. Por eso, voy a describir la estructura del Universo en cuatro dimensiones. Antes, no puedo dejar de reseñar los modelos de «multiversos», derivados de la teoría de supercuerdas, elegantes especulaciones físico-matemáticas de universos múltiples, en los que nuestro Universo tridimensional sería una proyección más de tres dimensiones, instalada en un espacio global de nueve.

A continuación intentaré hacer una descripción asequible de cómo imaginamos los astrónomos hoy el Universo, en el momento actual de su historia. Después, me detendré en alguna de las etapas más significativas de su evolución.

El Universo a gran escala se nos presenta hoy ante nuestros telescopios, con su poco más de 13.000 millones de años de vida, enormemente vacío. La materia aparece muy concentrada y jerarquizada gravitatoriamente en estrellas, con sus sistemas planetarios, en galaxias, en cúmulos de galaxias y en supercúmulos de galaxias (figura 4). Los enormes vacíos —planetarios, interestelares e intergalácticos— están llenos de materia muy diluida, pero que sumada representa la mayor parte de la materia ordinaria. La materia oscura también se acumula y ordena de manera análoga, al estar regida por la gravedad. Mientras, por el contrario, la energía oscura se reparte uniformemente por todos sitios.

Si fuésemos haciendo zoom, y acercándonos a cada rincón de nuestra propia galaxia —la Vía Láctea— nos encontraríamos con brillantes sistemas planetarios de uno o varios soles, con sus planetas, satélites, cometas y miríadas de cuerpos menores orbitando unos en torno a otros, y todos alrededor de sus respectivos centros de

Página anterior:

Figura 4. Galaxia conocida como Triángulo o M33, obtenida con el INT del Observatorio del Roque de los Muchachos del Instituto de Astrofísica de Canarias. En esta extraordinaria imagen de la menor de las tres galaxias espirales que conforman nuestro Grupo Local —compuesto además por Andrómeda, la Vía Láctea y unas cuantas más de menor tamaño— se pueden apreciar un sinnúmero de sus componentes: desde diversas clases de nubes interestelares, hasta estrellas en todas sus fases evolutivas, bellas nebulosas planetarias y remanentes de supernovas. (Créditos: IAC)

masas. Y como en cada galaxia puede haber cientos de miles de millones de ellos, algunos estarán naciendo en el seno de nubes de polvo y gas interestelares, entre convulsiones y reajustes, mientras otros estarían en sus etapas finales, implosionando y explotando, expulsando materia incandescente, partículas y radiación en un espectáculo dantesco y bellissimo. Mientras, la mayoría de los objetos galácticos los encontraremos en las etapas intermedias de sus respectivas peripecias vitales. La descripción de la vida y milagros de la multifacética y variopinta «fauna» galáctica nos ocuparía demasiadas páginas, llenas, eso sí, de colorido, belleza y drama. Otra cuestión, que sí nos importa y mucho, es la de la existencia de vida más allá de nuestro planeta. De ser la vida algo extendido en el Universo, tendría que estar influyendo, aunque no sepamos cómo aún, en su estructura. Por esta vía podrían venir reajustes conceptuales de nuestra cosmovisión aún mayores de los que se avecinan como consecuencia del Universo acelerado.

Los enormes enjambres de estrellas, gas, polvo y mucha materia oscura que son las galaxias no están aislados en el espacio, por el contrario los observamos fuertemente ligados entre sí por gravedad. Y se acoplan formando agrupaciones, cúmulos de galaxias, que a su vez se asocian en supercúmulos. Y estas desmesuradas acumulaciones de materia se nos presentan organizadas en mallas similares a una telaraña, con direcciones filamentosas privilegiadas de decenas de millones de años luz. Todo ello flotando en vacíos enormes.

No debemos olvidar que toda esta enormidad está en plena actividad y que todos los objetos celestes se están moviendo con velocidades alucinantes. Imaginar, por ello, un Universo mecánicamente regulado al modo de un reloj perfecto, es lo más contrario a la realidad. Las interacciones son múltiples y los choques frecuentes. Estos choques, ya sean entre las capas externas estelares y el medio circundante, ya sea entre nubes y supernubes interestelares o entre galaxias, resultan ser los mecanismos más eficaces para cincelar las galaxias y movilizar los entornos cósmicos (figura 1). La energía parece derrocharse en fenómenos violentísimos que observamos por doquier y que producen nuevos objetos celestes.

Embebida en toda esta superestructura, además, está la energía oscura contrarrestando con eficacia a la gravedad y expandiendo el espacio aceleradamente, generando sin duda acciones directas o indirectas en todos los escalones de la estructura cósmica. El que aún no las conozcamos, no quiere decir que no estén produciéndose.

Podemos quedarnos con la imagen simplificada de un Universo gigantesco, violento, en expansión acelerada, con la materia —la ordinaria de la que nosotros mismos estamos hechos, y la oscura— concentrada en islas plenas de acción empujada por la gravedad, empapada toda esta plétora por la energía oscura, y bañada en radiación electromagnética. Y en un rinconcito, nuestra pequeñísima Tierra, llena de vida, danzando en el espacio.

Después de este visionado semi-cinematográfico de cómo entendemos que debe ser la estructura actual del Universo, hay que decir algo de las etapas principales de su vida. Porque lo que ahora contemplamos, incluida la vida que bulle en el planeta Tierra, es consecuencia de su evolución general, determinada por unas leyes que tratamos de ir descubriendo. Precisamente conocer el nacimiento y evolución de todas y cada una de las partes del Cosmos es lo que subyace actualmente en toda investigación astronómica.

Evolución del Universo

Mucho se ha escrito sobre la flecha del tiempo, tratando de descubrir en qué dirección camina la evolución de nuestro Universo y, puesto que tuvo un principio, conocer cuál será su final.

Desde que el Universo dejó de ser inmutable, a principios del pasado siglo, queremos conocer su historia y, sobre todo, su evolución, pues en ella está la clave del origen de nuestra propia historia y de la presencia de vida inteligente en otros planetas de otros sistemas estelares. Pero una historia es un relato de acontecimientos en el tiempo, y por lo que parece, el tiempo, nuestro tiempo, empezó con el propio Universo al que pertenecemos y aún no sabemos a ciencia cierta cuál es la esencia real de este parámetro físico. Sin entrar en disquisiciones profundas, consideraré el tiempo como lo imaginamos de manera intuitiva: un continuo uniforme que se extiende desde la Gran Explosión hacia un futuro lejano.

La casi totalidad de la información que nos llega del mundo exterior está en forma de radiación electromagnética, y la primera instantánea retrospectiva del Universo nos la proporciona el Fondo Cósmico de Microondas. Para entonces, el Universo tenía unos 400.000 años, y ya antes le habían sucedido muchas cosas de gran trascendencia. Lo que pasó antes lo inferimos de los modelos cosmogónicos, el más aceptado de los cuales sigue siendo el llamado Modelo Estándar. No olvidemos que este modelo describe lo que sucedió después de la Gran Explosión, pero no aporta información sobre la singularidad en sí misma. Recordemos también que el modelo se desarrolló antes del descubrimiento de la expansión acelerada, y que los tres pilares sobre los que se sustenta son: la expansión decelerada, el Fondo Cósmico de Microondas y la nucleogénesis primordial, que produjo los primeros elementos ligeros que siguen dominando la materia. Su clave es que el Universo comenzó siendo muy caliente y muy denso, para al expandirse irse diluyendo y enfriando.

Para explicar la naturaleza y el nacimiento de la materia y sus interacciones, se siguen produciendo abundantes teorías y modelos muy especulativos, y de una complejidad sólo asequible a quienes a esto se dedican. Sus fundamentos empíricos, no obstante, son observaciones astronómicas y experimentos con los grandes aceleradores de partículas, que resultan aún insuficientes para poner luz entre tanta especulación físico-matemática. Digo todo esto para evi-

tar que se den por científicamente probadas la mayoría de las cosas desconcertantes que se afirman, incluidas las que voy a decir, sobre los primeros instantes del Universo.

Inmediatamente después de que todo empezase en la Gran Explosión, pasados tan sólo los primeros 10^{-35} segundos, cuando todas las fuerzas fundamentales estaban unificadas aún, el espacio sufrió una expansión exponencial prodigiosa —creció en un factor 10^{26} , en tan sólo 10^{-33} segundos—, según postulan los modelos inflacionarios, amparados en datos de la radiación de fondo. Tal expansión acelerada diluyó todo lo preexistente, suavizando las posibles variaciones en su densidad. Y aquí nos encontramos ya con la primera expansión acelerada, que supone algo tan insólito como que la densidad de energía debe ser positiva y mantenerse casi constante —este casi es muy importante— mientras su presión es negativa (figura 5). Y acaba con una brusca caída de la densidad. Obviamente, no sabemos cómo y por qué arranca y termina la inflación.

Durante el periodo inflacionario, la densidad de energía del espacio fluctúa mínimamente, debido a la naturaleza estadística de las leyes cuánticas que rigen a escalas subatómicas. Pero estas homogeneidades son agrandadas exponencialmente por la inflación, dando lugar a anisotropías en el Fondo Cósmico de Microondas, y son las semillas que marcan el devenir grandioso del Universo, son los embriones de las macroestructuras de galaxias y cúmulos de galaxias que hoy vemos. Sale el Universo de este periodo recalentado, y la energía potencial del vacío convertida en partículas calientes.

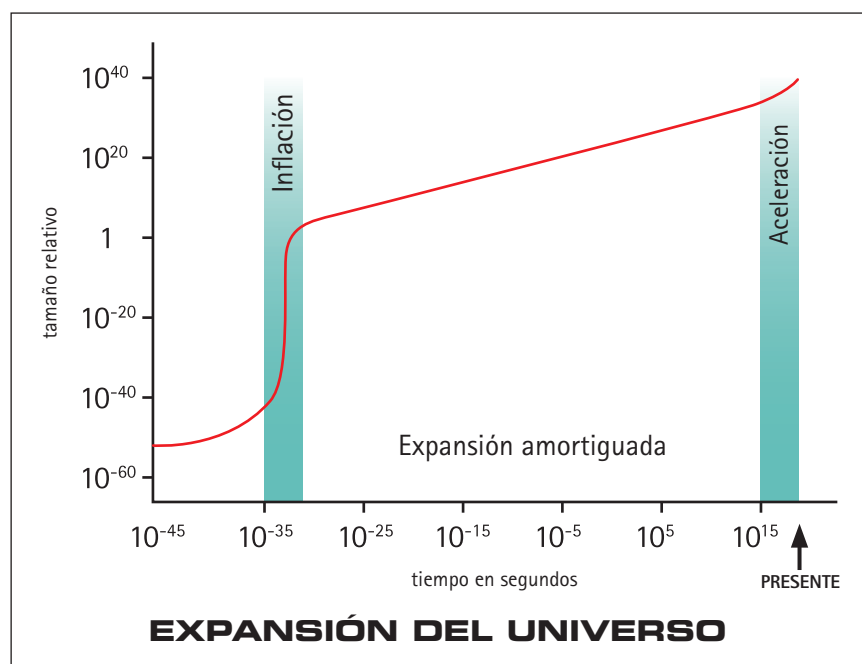


Figura 5. Aunque no es posible representar con realismo la expansión del espacio y mucho menos su extraordinaria «inflación», para formarnos una primera idea —aunque simplificada— de la expansión del Universo en función del tiempo, hemos dibujado esta gráfica. Se ha hecho en escalas logarítmicas para hacer visibles los detalles. Tanto «inflación» como «expansión acelerada» tienen en común una presión negativa que se contrapone a la atracción gravitatoria. La naturaleza de estos fenómenos nos es desconocida por el momento. (Créditos: IAC)

Siguiendo esta descripción sintética de la evolución del Universo que estoy haciendo de acuerdo con los modelos más aceptados, las diferentes partículas y antipartículas, junto con sus interacciones, van creándose al ritmo que la expansión y consiguiente enfriamiento del Universo lo van permitiendo. A solo 10^{-5} segundos ya hay bariones. Este «potaje» de partículas en continuo nacimiento y desaparición, siguió enfriándose y casi todas las partículas de materia y antimateria se aniquilaron entre sí. Pero, por razones que se desconocen, quedó un ligero exceso de bariones que no hallaron partículas de antimateria para aniquilarse y, por consiguiente, sobrevivieron a esta extinción.

Cuando la temperatura cayó a unos 3.000 grados, los protones y los electrones pudieron combinarse y formaron átomos, eléctricamente neutros, de hidrógeno. Se descola así la materia de la radiación, al dejar los fotones de interactuar con la materia de forma tan intensa, y la luz se desparrama por doquier. Estos fotones primigenios son los que constituyen la radiación de fondo de microondas. Para entonces el Universo tenía ya 400.000 años, y le habían pasado, como hemos visto, cosas muy importantes. Una de ellas fue la nucleosíntesis primordial, que determinó la preponderancia absoluta del hidrógeno y el helio en el Universo. Tal proceso, gobernado por la expansión, sólo pudo darse justo durante unos pocos minutos, por ello esta nucleosíntesis no pasó de los elementos más ligeros.

A continuación se abre una época gris, la que va de la liberación de la radiación de fondo de microondas al resurgimiento de la luz, con el nacimiento de las primeras galaxias y estrellas. Esta etapa del Universo nos es muy desconocida porque no hay radiaciones que observar, sin embargo resulta decisiva, pues en ella la gravedad se dedicó a ir ensamblando los objetos que hoy pueblan el Cosmos. Ese periodo concluye en los primeros millones de años, cuando la luz de las estrellas brotó en cantidad suficiente como para ionizar, con su radiación ultravioleta, el gas que ahora domina el espacio intergaláctico. Tales estrellas fueron muy peculiares: astros supermasivos, de cien masas solares o más y compuestos sólo de hidrógeno y helio. Todo esto está corroborado por los espectros observados de los cuásares, las galaxias y por las explosiones de rayos gamma más remotas, además de por el hallazgo de galaxias lejanas, a menos de mil millones de años de la Gran Explosión.

Se cree que una galaxia, o una estrella, empiezan a formarse cuando una región del espacio con mayor densidad que su entorno comienza a contraerse sobre sí misma, por efecto de su propia gravedad. Las galaxias, no se olvide, están formadas sobre todo por materia oscura que se nos escapa a la observación directa. Aunque la región esté sometida a la expansión global, el exceso de materia termina provocando su colapso, y así se crea un objeto ligado, ya sea una estrella, un cúmulo estelar o una galaxia. Son muchos, naturalmente, los añadidos y matices que son precisos para explicar lo que

sabemos al respecto, que ya es bastante, y abundantísimas son las investigaciones en curso sobre el origen y evolución de estrellas y galaxias. Sólo quería indicar cuál es el mecanismo básico de generación de la mayoría de los objetos que observamos.

Todo indica que, durante los primeros miles de millones de años del Universo, se produjeron frecuentes choques de galaxias, gigantescos brotes de formación estelar dentro de las mismas, y se generaron agujeros negros de más de miles de millones de masas solares. Fue una época extraordinariamente energética y agitada. Esta desahogada actividad parece estar ahora en declive, quizá como consecuencia de la aceleración en la expansión. En el Universo cercano sólo vemos tales prodigios de actividad en pequeñas galaxias, las grandes, tipo la Vía Láctea o Andrómeda, más reposadas, parecen haber entrado en etapa de madurez.

Aún no se puede precisar cuándo empezó a ser dominante la aceleración sobre la deceleración, aunque se ha apuntado que debió ser cuando el Universo tenía unos ocho mil millones de años. Anteriormente ya dijimos algo de lo que se especula sobre los efectos de la energía oscura. Sin duda, confirmada la realidad de las dos energías en competencia —la gravitatoria tratando de juntar la materia, y

la energía oscura tratando de separarla—, tienen que producirse modelos que expliquen todo lo que observamos. Para tener confianza en ellos, habrá que esperar a que los nuevos telescopios en tierra y en el espacio empiecen a producir datos significativos. En este sentido quiero decir que tenemos fuertes esperanzas puestas en el gran telescopio Canarias, que estará plenamente operativo en 2009 (figura 6). Los telescopios son las únicas máquinas del tiempo y cuanto mayores sean sus espejos, más profundamente penetran sus observaciones en el Cosmos y más retrocede lo que vemos en el tiempo. Con ellos pretendemos observar el nacimiento y la evolución de los objetos más primitivos.

Por lo mucho que nos toca, señalaremos que nuestro Sol y su Sistema Solar nacieron, cuando el Universo debería andar por los nueve mil millones de años, de una nube de material reciclado producido en el interior de estrellas precedentes que lo fueron arrojando al medio interestelar. Los elementos químicos de las moléculas orgánicas, que en nuestro planeta sustentan toda forma de vida, no se han podido crear *in situ*, necesariamente estaban ya en el disco protoplanetario, lo que nos permite decir con certeza que «somos polvo de estrellas». Lo cual va mucho más allá de una bella frase poética.

Hasta hace nada, el futuro del Universo se predecía en base a la energía gravitatoria y a la termodinámica, en función de la cantidad de materia que contuviese. Si la masa estaba por encima del valor crítico calculado de acuerdo con los modelos al uso, se vaticinaba que iría frenándose la expansión hasta colapsar en una nueva singularidad y rebotar en un proceso oscilante de Grandes Explosiones y sus consiguientes colapsos. De lo contrario, se iría expandiendo indefinidamente. En estos momentos hay que contar con la aceleración del Universo, y ello nos lleva a imaginar un futuro diferente.

Final

De continuar la expansión acelerada, las galaxias, empezando por las más lejanas, habrán ido desapareciendo de nuestra vista detrás del horizonte de sucesos. Dentro de unos miles de millones de años las galaxias cercanas se habrán fusionado formando un conjunto de estrellas gigantesco ligado por gravedad, una mega súper galaxia —mesuga— envuelta en un espacio oscuro y vacío. La radiación de fondo de microondas estará tan diluida que será indetectable. Mucho antes de que todo se enfríe y se acabe, quedaremos aislados en el espacio y nuestro Universo asequible será sólo nuestra propia mesuga.

Desolador final de nuestro mundo físico. Pero ¿es realmente el final? ¿Cómo evolucionarán internamente las miríadas de millones de mesugas al quedar aisladas? ¿Habrá algún mecanismo que las conecte de alguna forma, pese a que la separación entre ellas siga aumentando aceleradamente?

Está visto que ¡nunca perdemos la esperanza, ni nuestro afán de eternidad! Así somos. Estamos, además, determinados para ser curiosos, tenemos el instinto del saber.



Figura 6. Vista exterior del Gran Telescopio CANARIAS –GTC–, situado en el Observatorio del Roque de los Muchachos del Instituto de Astrofísica de Canarias. Es el mayor y más avanzado telescopio óptico-infrarrojo del mundo, con un espejo primario segmentado de 10,4 metros de diámetro. La baja nubosidad de este Observatorio —las nubes que se ven están por debajo, aunque no lo parezca por la perspectiva de la imagen— junto con la transparencia y estabilidad de su atmósfera, hacen de este sitio uno de los escasísimos lugares del planeta aptos para instalar los más avanzados telescopios. El GTC, con su gran objetivo, su excelente óptica y su potente instrumentación focal, más la extraordinaria calidad astronómica del cielo de La Palma, constituye una poderosísima herramienta para penetrar en los secretos del Universo. (Créditos: IAC)

En los *sapiens-sapiens* parece ser un impulso tan básico como el de reproducirse. ¿Tendrá esto algo que ver con la expansión acelerada de nuestro Universo?

Tal y como están las cosas de momento, parece que ha hecho falta la existencia conjunta de las energías antagónicas, gravitatoria y oscura, para que se formase el Universo, y dentro de él nuestro Sol y nuestra Tierra, y que tras laboriosos procesos evolutivos previos, nuestros padres pudiesen engendrarnos. Habría bastado con que la energía oscura fuese un poco más débil o un poco más poderosa para que no hubiésemos existido.

Tanto si descendemos a las minimeces de las últimas partículas subnucleares, como si nos perdemos en las inmensidades del Cosmos, todo lo encontramos pleno de actividad, movido por fuerzas poderosas. Lo mismo podemos decir de la vida en todas sus facetas, ya se manifieste en organismos unicelulares o en bosques impenetrables, también aquí la actividad es incesante. Debe ser algo consustancial con nuestro Universo: nada es estático, todo es acción y evolución. La energía abunda y parece derrocharse: en las galaxias y en las estrellas son frecuentes los episodios cataclísmicos, de una potencia desmesurada. Toda

esta agitación, frecuentemente catastrófica, es la manifestación de lo que podríamos denominar «ímpetu cósmico», que lo llena todo y lo empuja todo. Este «ímpetu» se expresa como energías, que en un intento sistematizador hasta se pondrían catalogar en dos grandes grupos: las energías físicas y las energías vitales. En algún momento habrá que conceptualizar todo esto y darle formalismo matemático.

Quienes hayan disfrutado de noches estrelladas en plena naturaleza, lejos de la polución y el bullicio de las ciudades, y se hayan dejado llevar por las sensaciones, los sentimientos y la imaginación, en ese estado de gracia, habrán sentido la fuerza profunda de las preguntas fundamentales. Aunque ninguna de ellas haya sido respondida en lo escrito anteriormente, si han leído atentamente habrán sentido brillar la chispa de la inteligencia humana enfrentándose con éxito a la desmesura del Universo y a sus atractivos misterios. La verdad es que, aunque nuestro desconocimiento tienda a infinito, lo ya conocido es mucho y supone un triunfo notable que dignifica a toda la humanidad. Mientras esta especie increíble, malvada y cerril, a la par que amorosa e inteligente, siga existiendo, continuará mirando e interrogando al cielo.

Bibliografía

- Adams, F. C. y G. Laughlin. *The five ages of the universe: inside the physics of eternity*. Free Press, 2000.
- Caldwell, R. R. «Dark energy». *Physics World*, 1.17 / 5 (2004): 37-42.
- , M. Hamionkowski y N. N. Weinberg. «The phantom energy and cosmic doomsday». *Physical Review Letters*, 91 (2003) 071301.
- Dine, M. y A. Kusenko. «Origin of the matter antimatter asymmetry». *Reviews of Modern Physics*, 76 (2004): 1-30.
- Freedman, W. L. y M. Turner, Michael. «Measuring and understanding the universe». *Reviews of Modern Physics*, 75 (2003): 1.433-1.447.
- Harrison, E. R. *Cosmology: the science of the universe*. Cambridge University Press, 2000.
- Hertzberg, M., R. Tegmark, M. Shamit Kachru, J. Shelton y O. Ozcan. «Searching for inflation in simple string theory models: an astrophysical perspective». *Physical Review*, 76 / 103521 (2007): 37-45.
- Kirshner, R. P. *The extravagant universe: exploding stars, dark energy, and accelerating cosmos*. Princeton University Press, 2004.
- Krauss, L. y G. Starkman. «Life, the universe and nothing: life and death in an everexpanding universe». *Astrophysical Journal*, 531 / 22 (2000): 22-30.
- Peebles, P. J. E. y B. Ratra. «The cosmological constant and dark energy». *Reviews of Modern Physics*, 75 (2003): 559-606.
- Primack, J. y N. E. Abrams. *The View from the Centre of the Universe: Discovering Our Extraordinary Place in the Cosmos*. Riverhead, 2006.
- Silk, J. *The infinite cosmos: questions from the frontiers of cosmology*. Oxford University Press, 2006.
- Srianand, T. A., P. Petitjean y C. Ledoux. «The cosmic microwave background radiation temperature at a redshift of 2.34». *Nature*, 408 / 6815 (2000): 931-935.
- Wilson, G. *et al.* «Star formation history since $z = 1$ as inferred from rest-frame ultraviolet luminosity density evolution». *Astronomical Journal*, 124 (2002): 1.258-1.265.

el mundo después de la revolución: la física de la segunda mitad del siglo xx

JOSÉ MANUEL SÁNCHEZ RON

Las grandes revoluciones del siglo xx

Durante la primera mitad del siglo xx –estrictamente en su primer cuarto– se produjeron dos grandes revoluciones científicas. Fue en la física donde tuvieron lugar tales cataclismos cognitivos, a los que conocemos bajo la denominación de revoluciones relativista y cuántica, asociadas a la formulación de las teorías especial y general de la relatividad (Einstein 1905a 1915) y de la mecánica cuántica (Heisenberg 1925; Schrödinger 1926).

Relatividad

Mucho se ha escrito y escribirá en el futuro sobre la importancia de estas formulaciones teóricas y cómo afectaron al conjunto de la física antes incluso de que la centuria llegase a su mitad. Creada para resolver la «falta de entendimiento» que crecientemente se percibía entre la mecánica newtoniana y la electrodinámica de James Clerk Maxwell (1831-1879), la teoría de la relatividad especial obligó a modificar radicalmente las ideas y definiciones –vigentes desde que Isaac Newton (1642-1727) las incorporase al majestuoso edificio contenido en su *Philosophiae Naturalis Principia Mathematica* (1687)– de conceptos tan básicos desde el punto de vista físico, ontológico y epistemológico como son espacio, tiempo y materia (masa). El resultado, en el que las medidas de espacio y tiempo dependían del estado de movimiento del observador y la masa, m , era equivalente a la energía, E (la célebre expresi-

ón $E=m\cdot c^2$, donde c representa la velocidad de la luz), abrió nuevas puertas a la comprensión del mundo físico; sirvió, por ejemplo, para comenzar a entender cómo era posible que los elementos radiactivos (uranio, polonio, radio, torio) que Henri Becquerel (1852-1908) junto a Marie (1867-1934) y Pierre Curie (1859-1906) habían sido los primeros en estudiar (1896, 1898), emitiesen radiaciones de manera continua, sin aparentemente perder masa.

¡Y qué decir de la teoría general de la relatividad, que explicaba la gravedad a costa de convertir el espacio –mejor dicho, el cuatridimensional espacio-tiempo– en curvo y con una geometría variable! Inmediatamente se comprobó que con la nueva teoría einsteiniana era posible comprender mejor que con la gravitación universal newtoniana los fenómenos perceptibles en el Sistema Solar (se resolvió, por ejemplo, una centenaria anomalía en el movimiento del perihelio de Mercurio). Y por si fuera poco, enseguida el propio Einstein (1917) tuvo la osadía intelectual de aplicar la teoría de la relatividad general al conjunto del Universo, creando así la cosmología como disciplina auténticamente científica, predictiva. Es cierto que el modelo que Einstein propuso entonces, uno en el que el Universo era estático, no sobrevivió finalmente, pero lo importante, abrir la puerta al tratamiento científico del universo, constituyó un acontecimiento difícilmente igualable en la historia de la ciencia.¹

Para encontrar la solución exacta de las ecuaciones de la cosmología relativista que utilizó, Einstein (1879-1955)

1

Para poder construir un modelo de universo estático, Einstein tuvo que modificar las ecuaciones básicas de la relatividad general, incluyendo un término adicional, en el que aparecía una «constante cosmológica».

se guio por consideraciones físicas. Otros matemáticos o físicos con especiales sensibilidades y habilidades matemáticas, no siguieron semejante senda, hallando muy pronto nuevas soluciones exactas –que implícitamente representaban otros modelos de universo– recurriendo únicamente a técnicas matemáticas para tratar las complejas (un sistema de diez ecuaciones no lineales en derivadas parciales) ecuaciones de la cosmología relativista. Así, Alexander Friedmann (1888-1925), Howard Robertson (1903-1961) y Arthur Walker (n. 1909) encontraron soluciones que implicaban modelos de universo en expansión. De hecho, hubo otro científico que obtuvo un resultado similar: el sacerdote católico belga Georges Lemaître (1894-1966), pero éste debe ser mencionado por separado ya que al igual que había hecho Einstein con su modelo estático, Lemaître (1927) se basó en consideraciones físicas para defender la idea de una posible, real, expansión del Universo.

Ahora bien, todos estos modelos surgían de soluciones de las ecuaciones cosmológicas; esto es, se trataba de posibilidades teóricas. La cuestión de cómo es realmente el Universo –¿estático?, ¿en expansión?– quedaba aún por dilucidar, para lo cual el único juez aceptable era la observación.

La gloria imperecedera de haber encontrado evidencia experimental a favor de que el Universo se expande pertenece al astrofísico estadounidense Edwin Hubble (1889-1953), quien se benefició del magnífico telescopio reflector con un espejo de 2,5 metros de diámetro que existía en el observatorio de Monte Wilson (California) en el que trabajaba, al igual que de unos excelentes indicadores de distancia, las cefeidas, estrellas de luminosidad



2
Hoyle (1948), Bondi y Gold (1948).

Edwin Hubble tomando fotografías con el telescopio de 2,5 metros de Monte Wilson (1924).

variable en las que se verifica una relación lineal entre la luminosidad intrínseca y el periodo de cómo varía esa luminosidad (Hubble 1929; Hubble y Humason 1931). Y si, como Hubble sostuvo, el Universo se expandía, esto quería decir que debió existir en el pasado (estimado inicialmente en unos diez mil millones de años, más tarde en quince mil millones y en la actualidad en unos trece mil setecientos millones) un momento en el que toda la materia habría estado concentrada en una pequeña extensión: el «átomo primitivo» de Lemaître, o, una idea que tuvo más éxito, el Big Bang (Gran Estallido).

Nació así una visión del Universo que en la actualidad forma parte de la cultura más básica. No fue, sin embargo, siempre así. De hecho, en 1948, cuando terminaba la primera mitad del siglo, tres físicos y cosmólogos instalados en Cambridge: Fred Hoyle (1915-2001), por un lado, y Hermann Bondi (1919-2005) y Thomas Gold (1920-2004), por otro (los tres habían discutido sus ideas con anterioridad a la publicación de sus respectivos artículos), dieron a conocer un modelo diferente del Universo en expansión: la cosmología del estado estable, que sostenía que el Universo siempre ha tenido y tendrá la misma forma (incluyendo densidad de materia, lo que, debido a la evidencia de la expansión del Universo, obligaba a introducir la creación de materia para que un «volumen» de Universo tuviese siempre el mismo contenido aunque estuviese dilatándose); en otras palabras: que el Universo no tuvo ni un principio ni tendrá un final.²

A pesar de lo que hoy podamos pensar, imbuidos como estamos en «el paradigma del Big Bang», la cosmología del estado estable ejerció una gran influencia durante la década de 1950. Veremos que fue en la segunda mitad del siglo cuando finalmente fue desterrada (salvo para unos pocos fieles, liderados por el propio Hoyle).

Física cuántica

La segunda gran revolución a la que hacía referencia es la de la física cuántica. Aunque no es rigurosamente exacto, hay sobrados argumentos para considerar que el punto de partida de esta revolución tuvo lugar en 1900, cuando mientras estudiaba la distribución de energía en la radiación de un cuerpo negro, el físico alemán Max Planck (1858-1947) introdujo la ecuación $E=h\nu$ donde E es, como en el caso de expresión relativista, la energía, h una constante universal (denominada posteriormente «constante de Planck») y ν la frecuencia de la radiación involucrada (Planck 1900). Aunque él se resistió de entrada a apoyar la idea de que este resultado significaba que de alguna manera la radiación electromagnética (esto es, la luz, una onda *continua* como se suponía hasta entonces) se podía considerar también como formada por «corpúsculos» (posteriormente denominados «fotones») de energía $h\nu$, semejante implicación terminó imponiéndose, siendo en este sentido Einstein (1905b) decisivo. Se trataba de la «dualidad onda-corpúsculo».



Werner Heisenberg en Gotinga (hacia 1924).

Durante un cuarto de siglo, los físicos pugnaron por dar sentido a los fenómenos cuánticos, entre los que terminaron integrándose también la radiactividad, la espectroscopia y la física atómica. No es posible aquí ofrecer ni siquiera un esbozo del número de científicos que trabajaron en este campo, de las ideas que manejaron y los conceptos que introdujeron, ni de las observaciones y experimentos realizados. Únicamente puedo decir que un momento decisivo en la historia de la física cuántica se produjo en 1925, cuando un joven físico alemán de nombre Werner Heisenberg (1901-1976) desarrolló la primera formulación coherente de una mecánica cuántica: la *mecánica cuántica matricial*. Poco después, en 1926, el austriaco Erwin Schrödinger (1887-1961) encontraba una nueva versión (pronto se comprobó que ambas eran equivalentes): la *mecánica cuántica ondulatoria*.

Si la exigencia de la constancia de velocidad de la luz contenida en uno de los dos axiomas de la teoría de la relatividad especial, la dependencia de las medidas espaciales y temporales del movimiento del observador o la curvatura dinámica del espacio-tiempo constituían resultados no sólo innovadores sino sorprendentes, que violentan nuestro «sentido común», mucho más chocantes resultaron ser aquellos contenidos o deducidos en la mecánica cuántica, de los que es obligado recordar al menos dos: (1) la interpretación de la función de onda de la ecuación de Schrödinger debida a Max Born (1882-1970), según la cual tal función –el elemento básico en

la física cuántica para describir el fenómeno considerado– representa la probabilidad de que se dé un resultado concreto (Born 1926); y (2) el principio de incertidumbre (Heisenberg 1927), que sostiene que magnitudes canónicamente conjugadas (como la posición y la velocidad, o la energía y el tiempo) sólo se pueden determinar simultáneamente con una indeterminación característica (la constante de Planck): $\Delta x \Delta p \geq h$, donde x representa la posición y p el momento lineal (el producto de la masa por la velocidad). A partir de este resultado, al final de su artículo Heisenberg extraía una conclusión con implicaciones filosóficas de largo alcance: «En la formulación fuerte de la ley causal "Si conocemos exactamente el presente, podemos predecir el futuro", no es la conclusión, sino más bien la premisa la que es falsa. *No podemos conocer*, por cuestiones de principio, el presente en todos sus detalles». Y añadía: «En vista de la íntima relación entre el carácter estadístico de la teoría cuántica y la imprecisión de toda percepción, se puede sugerir que detrás del universo estadístico de la percepción se esconde un mundo "real" regido por la causalidad. Tales especulaciones nos parecen –y hacemos hincapié en esto– inútiles y sin sentido. Ya que la física tiene que limitarse a la descripción formal de las relaciones entre percepciones».

La mecánica cuántica de Heisenberg y Schrödinger abrió un mundo nuevo, científico al igual que tecnológico, pero no era en realidad sino el primer paso. Existían aún muchos retos pendientes, como, por ejemplo, hacerla compatible con los requisitos de la teoría de la relatividad especial, o construir una teoría del electromagnetismo, una electrodinámica, que incorporase los requisitos cuánticos. Si Einstein había enseñado, y la física cuántica posterior incorporado en su seno, que la luz, una onda electromagnética, estaba cuantizada, esto es, que al mismo tiempo que una onda también era una «corriente» de fotones, y si la electrodinámica que Maxwell había construido en el siglo XIX describía la luz únicamente como una onda, sin ninguna relación con la constante de Planck, entonces era evidente que algo fallaba, que también había que cuantizar el campo electromagnético.

No fue necesario, sin embargo, esperar a la segunda mitad del siglo XX para contar con una electrodinámica cuántica. Tal teoría, que describe la interacción de partículas cargadas mediante su interacción con fotones, fue construida en la década de 1940, de manera independiente, por un físico japonés y dos estadounidenses: Sin-itiro Tomonaga (1906-1979), Julian Schwinger (1918-1984) y Richard Feynman (1918-1988).³

La electrodinámica cuántica representó un avance teórico considerable, pero tampoco significaba, ni mucho menos, el final de la historia cuántica; si acaso, ascender un nuevo peldaño de una escalera cuyo final quedaba muy lejos. En primer lugar porque cuando la teoría de Tomonaga-Schwinger-Feynman fue desarrollada ya estaba claro que además de las tradicionales fuerzas electromagnética y gravitacio-

3

Fukuda, Miyamoto y Tomonaga (1949), Schwinger (1949) y Feynman (1949). Por sus trabajos los tres compartieron el Premio Nobel de Física en 1965.

nal existen otras dos: la débil, responsable de la existencia de la radiactividad, y la fuerte, que unía a los constituyentes (protones y neutrones) de los núcleos atómicos.⁴ Por consiguiente, no bastaba con tener una teoría cuántica de la interacción electromagnética, hacía falta además construir teorías cuánticas de las tres restantes fuerzas.

Relacionado íntimamente con este problema, estaba la proliferación de partículas «elementales». El electrón fue descubierto, como componente universal de la materia, en 1897 por Joseph John Thomson (1856-1940). El protón (que coincide con el núcleo del hidrógeno) fue identificado definitivamente gracias a experimentos realizados en 1898 por Wilhelm Wien (1864-1928) y en 1910 por Thomson. El neutrón (partícula sin carga) fue descubierto en 1932 por el físico inglés James Chadwick (1891-1974). Y en diciembre de este año, el estadounidense Carl Anderson (1905-1991) hallaba el positrón (idéntico al electrón salvo en que su carga es opuesta, esto es, positiva), que ya había sido previsto teóricamente en la ecuación relativista del electrón, introducida en 1928 por uno de los pioneros en el establecimiento de la estructura básica de la mecánica cuántica, el físico inglés Paul Dirac (1902-1984).

Electrones, protones, neutrones, fotones y positrones no serían sino los primeros miembros de una extensa familia (mejor, familias) que no hizo más que crecer desde entonces, especialmente tras la entrada en funcionamiento de unas máquinas denominadas «aceleradores de partículas». En el establecimiento de esta rama de la física, el ejemplo más característico de lo que se ha venido en llamar *Big Science* (Gran Ciencia), ciencia que requiere de enormes recursos económicos y de equipos muy numerosos de científicos y técnicos, nadie se distinguió más que Ernest O. Lawrence (1901-1958), quien a partir de la década de 1930 desarrolló en la Universidad de Berkeley (California) un tipo

de esos aceleradores, denominados «ciclotrones», en los que las partículas «elementales» se hacían girar una y otra vez, ganando en cada vuelta energía, hasta hacerlas chocar entre sí, choques que se fotografiaban para luego estudiar sus productos, en los que aparecían nuevas partículas «elementales». Pero de esta rama de la física, denominada «de altas energías», volveré a hablar más adelante, cuando trate de la segunda mitad del siglo xx; ahora basta con decir que su origen se encuentra en la primera mitad de esa centuria.

Establecido el marco general, es hora de pasar a la segunda mitad del siglo, a la que está dedicada el presente artículo. Y comenzaré por el escenario más general: el Universo, en el que la interacción gravitacional desempeña un papel central, aunque, como veremos, no exclusivo, particularmente en los primeros instantes de su existencia.

El mundo de la gravitación

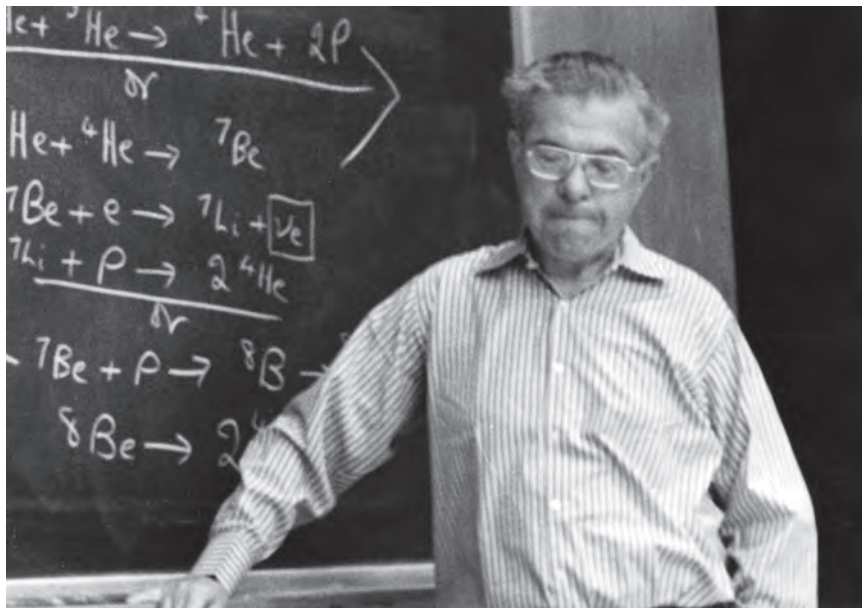
Evidencias de la expansión del Universo: la radiación cósmica de microondas

Señalé antes que no todos los físicos, astrofísicos y cosmólogos entendieron la expansión descubierta por Hubble como evidencia de que el Universo tuvo un comienzo, un Big Bang. La cosmología del estado estable de Hoyle-Bondi-Gold proporcionaba un marco teórico en el que el Universo había sido siempre igual, y esta idea fue bien aceptada por muchos. Sin embargo, en la década siguiente a la de su formulación, la de 1950, comenzó a tener problemas. El que fuese así se debió no a consideraciones teóricas, sino a las nuevas posibilidades observacionales que llegaron de la mano del desarrollo tecnológico. Es éste un punto que merece la pena resaltar: eso que llamamos ciencia es producto de una delicada combinación entre teoría y observación. No hay, efectivamente, ciencia sin la construcción de sistemas (teorías) que describen conjuntos de fenómenos, pero mucho menos la hay sin observar lo que realmente sucede en la naturaleza (simplemente, no somos capaces de imaginar cómo se comporta la naturaleza). Y para observar se necesitan instrumentos; cuanto más poderosos –esto es, capaces de mejorar las potencialidades de nuestros sentidos–, mejor. Y esto equivale a desarrollo tecnológico.

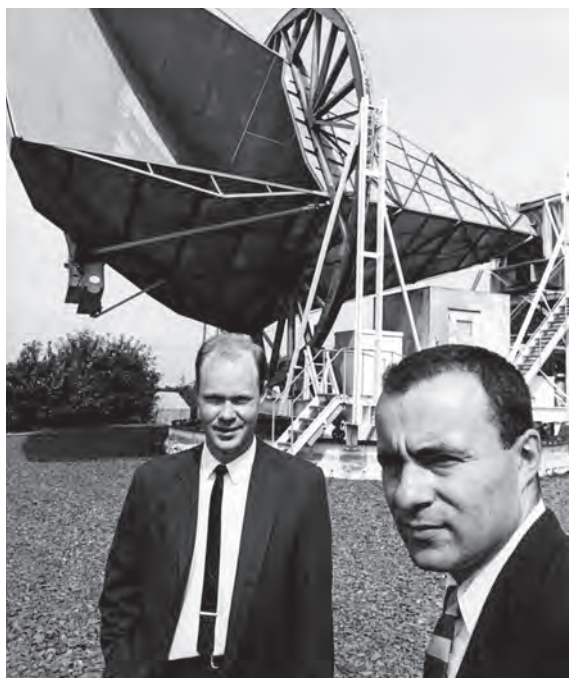
Sucede que la segunda mitad del siglo xx fue una época en la que la tecnología experimentó un desarrollo gigantesco, mucho mayor que en el pasado, y esto repercutió muy positivamente en el avance de la ciencia, en general, y de la astrofísica y cosmología en particular. En lo relativo a los problemas que afectaron a la cosmología del estado estable, a los que antes me refería, tales dificultades nacieron del desarrollo de la radioastronomía, una disciplina que había dado sus primeros pasos en la década de 1930, gracias a los trabajos de Karl Jansky (1905-1950), un ingeniero eléctrico que trabajaba para los laboratorios Bell (estrictamente Bell Telephone Laboratories), el «departamento» de American Telephone and Telegraph Corpora-

4

Hasta entonces se suponía que los núcleos atómicos estaban constituidos por protones (cargados positivamente) y electrones (cargados negativamente). ¿Cómo si no, se pensaba, podía explicarse la emisión de electrones (radiación beta) que tenía lugar en los procesos radiactivos? La desintegración beta terminaría siendo explicada utilizando una de las propiedades más llamativas de la física cuántica: la creación y aniquilación de partículas: los electrones no están en el núcleo, se crean en la desintegración beta.



Fred Hoyle en un seminario sobre nucleosíntesis en la Universidad Rice (Estados Unidos, marzo de 1975).



Robert Wilson y Arno Penzias delante de la antena con la que descubrieron el fondo cósmico de microondas.

tion encargado de la investigación y el desarrollo. En 1932, mientras buscaba posibles fuentes de ruido en emisiones de radio, Jansky detectó emisiones eléctricas procedentes del centro de nuestra galaxia. A pesar de la importancia que visto retrospectivamente asignamos ahora a tales observaciones, Jansky no continuó explorando las posibilidades que había abierto; al fin y al cabo, el mundo de la investigación fundamental no era el suyo.

No inmediatamente, pero sí pronto aquellas antenas primitivas se convirtieron en refinados radiotelescopios; habitualmente discos de cada vez mayor diámetro, que recogían radiación electromagnética procedente del espacio. La importancia de estos instrumentos para el estudio del Universo es obvia: los telescopios ópticos en los que se basaba hasta entonces la astrofísica únicamente estudiaban un rango muy pequeño del espectro electromagnético; eran, por así decir, casi «ciegos».

Uno de los primeros lugares en los que floreció institucionalmente la radioastronomía fue en Cambridge (Inglaterra). Fue allí donde Martin Ryle (1918-1984), continuó decididamente por la senda esbozada por Jansky. En semejante tarea se vio ayudado por los conocimientos que había obtenido durante la Segunda Guerra Mundial (trabajó entonces en el Telecommunications Research Establishment gubernamental, más tarde bautizado como Royal Radar Establishment), así como por la mejora que esa conflagración había significado para la instrumentación electrónica. Utilizando radiotelescopios, algunos de cuyos componentes diseñó él mismo, Ryle identificó en 1950 cincuenta radiofuentes, número que aumentó radicalmente cinco años más tarde, cuando llegó a las dos mil. Uno de sus hallazgos fue

descubrir una radio-fuente en la constelación de Cygnus, situada a 500 años-luz de la Vía Láctea. Al ver más lejos en el espacio, estaba viendo también más atrás en el tiempo (las señales que recibía habían sido emitidas hacía mucho, el tiempo que les había costado llegar a la Tierra). Con sus observaciones se estaba, por tanto, adentrando en la historia pasada del universo. Hubble había dado el primer gran paso en el camino de la cosmología observacional; Ryle – que recibió el Premio Nobel de Física en 1974– el segundo.

Gracias a sus observaciones con radio-fuentes, Ryle llegó a conclusiones que iban en contra de la cosmología del estado estable, reivindicando así la del Big Bang. Al analizar las curvas que relacionaban el número de radioestrellas por unidad de ángulo sólido con la intensidad que emiten, Ryle (1955) concluía que no veía la manera en la que las observaciones se pudiesen explicar en términos de la teoría del estado estable».

Mucho más concluyente a favor de la existencia en el pasado de un gran estallido fue otro descubrimiento, uno de los más célebres e importantes en toda la historia de la astrofísica y cosmología: el del fondo de radiación de microondas.

En 1961, E. A. Ohm, un físico de una de las instalaciones de los laboratorios Bell, situada en Crawford Hill, New Jersey, construyó un radiómetro para recibir señales de microondas procedentes del globo Echo (un reflector de señales electromagnéticas lanzado en 1960) de la NASA. No era una casualidad: los laboratorios Bell querían comenzar a trabajar en el campo de los satélites de comunicación. En observaciones realizadas en la longitud de onda de 11 cm., Ohm encontró un exceso de temperatura de 3,3°K (grados kelvin) en la antena, pero este resultado apenas atrajo alguna atención.⁵

Otro de los instrumentos que se desarrollaron por entonces en Crawford Hill fue una antena en forma de cuerno, una geometría que reducía las interferencias. El propósito inicial era utilizarla para comunicarse, vía el globo Echo, con el satélite Telstar de la compañía (la antena debía ser muy precisa, ya que debido a la forma del globo, las señales que incidiesen en él se difundirían mucho). En 1963, sabiendo de la existencia de esta antena, Robert Wilson (n. 1936) abandonó su puesto posdoctoral en el Instituto Tecnológico de California (Caltech) para aceptar un trabajo en los laboratorios Bell. Arno Penzias (n. 1933), un graduado de la Universidad de Columbia (Nueva York) tres años mayor que Wilson, ya llevaba por entonces dos años en los laboratorios. Afortunadamente, aquel mismo año la antena, pequeña pero de gran sensibilidad, pudo ser utilizada para estudios de radioastronomía, ya que la compañía decidió abandonar el negocio de comunicaciones vía satélite. Realizando medidas para una longitud de onda de 7,4 centímetros, Penzias y Wilson encontraron una temperatura de 7,5°K, cuando debía haber sido únicamente de 3,3°K. Además, esta radiación (o temperatura) suplementaria, que se creía efecto de algún ruido de fondo, resultó ser independiente de la

5

La relación entre temperaturas y longitudes de onda se puede obtener a partir de leyes como la de Stefan-Boltzmann, Wien o Planck. La relación entre grados celsius y kelvin viene definida por las relaciones siguientes: 0°C equivale a 273,15°K y 100°C a 373,15°K.

dirección en la que se dirigiese la antena. Los datos obtenidos indicaban que lo que estaban midiendo no tenía origen ni atmosférico, ni solar, ni galáctico. Era un misterio.

Después de descartar que los ruidos proviniesen de la propia antena, la única conclusión posible era que tenía algo que ver con el cosmos, aunque no se sabía cuál podía ser la causa. La respuesta a esta cuestión llegó de algunos colegas de la cercana Universidad de Princeton, algunos de los cuales, como James Peebles (n. 1935), ya habían considerado la idea de que si hubo un Big Bang debería existir un fondo de ruido remanente del universo primitivo, un ruido que, en forma de radiación, correspondería a una temperatura mucho más fría (debido al enfriamiento asociado a la expansión del Universo) que la enorme que debió producirse en aquella gran explosión. Las ideas de Peebles habían animado a su colega en Princeton Robert Dicke (1916-1995) a iniciar un experimento destinado a encontrar esa radiación de fondo cósmico, tarea en la que se les adelantaron, sin pretenderlo, Penzias y Wilson. Aun así, fue el grupo de Princeton el que suministró la interpretación de las observaciones de Penzias y Wilson (1965), que éstos publicaron sin hacer ninguna mención a sus posibles implicaciones cosmológicas. La temperatura correspondiente a esa radiación situada en el dominio de las microondas corresponde según las estimaciones actuales a unos 2,7°K (en su artículo de 1965, Penzias y Wilson daban un valor de 3,5°K).

El que Penzias y Wilson detectasen el fondo de radiación de microondas en un centro dedicado a la investigación industrial, en donde se disponía y desarrollaban nuevos instrumentos es significativo. Expresa perfectamente la ya mencionada necesidad de instrumentos más precisos, de nueva tecnología, para avanzar en el conocimiento del Universo. Según se dispuso de esta tecnología, fue ampliándose la imagen del cosmos. Y así llegaron otros descubrimientos, de los que destacaré dos: púlsares y cuásares.

Púlsares y cuásares

En 1963, un radioastrónomo inglés, Cyril Hazard, que trabajaba en Australia, estableció con precisión la posición de una poderosa radio-fuente, denominada 3C273. Con estos datos, el astrónomo holandés Maarten Schmidt (n. 1929), del observatorio de Monte Palomar (California), localizó ópticamente el correspondiente emisor, encontrando que las líneas del espectro de 3C273 estaban desplazadas hacia el extremo del rojo en una magnitud que revelaba que se alejaba de la Tierra a una velocidad enorme: 16% de la velocidad de la luz. Utilizando la ley de Hubble, que afirma que la distancia de las galaxias entre sí son directamente proporcionales a su velocidad de recesión, se deducía que 3C273 estaba muy alejada, lo que a su vez implicaba que se trataba de un objeto extremadamente luminoso, más de cien veces que una galaxia típica. Fueron bautizados como *quasi-stellar sources* (fuentes casi-estelares), esto es, *quasars* (cuásares), y se piensa que se trata de galaxias con núcleos muy activos.

Desde su descubrimiento, se han observado varios millones de cuásares, aproximadamente el 10% del número total de galaxias brillantes (muchos astrofísicos piensan que una buena parte de las galaxias más brillantes pasan durante un breve periodo por una fase en la que son cuásares). La mayoría están muy alejados de nuestra galaxia, lo que significa que la luz que se ve ha sido emitida cuando el universo era mucho más joven. Constituyen, por consiguiente, magníficos instrumentos para el estudio de la historia del Universo.

En 1967, Jocelyn S. Bell (n. 1943), Anthony Hewish (n. 1924) y los colaboradores de éste en Cambridge construyeron un detector para observar cuásares en las frecuencias radio. Mientras lo utilizaba, Bell observó una señal que aparecía y desaparecía con gran rapidez y regularidad. Tan constante era el periodo que parecía tener un origen artificial (¿acaso una fuente extraterrestre inteligente?). No obstante, tras una cuidadosa búsqueda Bell y Hewish concluyeron que estos «púlsares», como finalmente fueron denominados, tenían un origen astronómico (Hewish, Bell, Pilkington, Scott y Collins 1968).⁶ Ahora bien, ¿qué eran estas radio-fuentes tan regulares? La interpretación teórica llegó poco después, de la mano de Thomas Gold, uno de los «padres» de la cosmología del estado estable, reconvertido ya al modelo del Big Bang. Gold (1968) se dio cuenta de que los periodos tan pequeños implicados (del orden de 1 o 3 segundos en los primeros púlsares detectados) exigían una fuente de tamaño muy pequeño. Las enanas blancas eran demasiado grandes para rotar o vibrar con tal frecuencia, pero no así las estrellas de neutrones.⁷ Pero ¿el origen de las señales recibidas se debía a vibraciones o a rotaciones de estas estrellas? No a vibraciones, porque en estrellas de neutrones éstas eran demasiado elevadas (alrededor de mil veces por segundo) para explicar los periodos de la mayoría de los pulsares. Por consiguiente, los púlsares tenían que ser estrellas de neutrones en rotación. En la actualidad, cuando se han descubierto púlsares que emiten rayos X o gamma (incluso algunos luz en el espectro óptico), también se admiten otros mecanismos para la producción de la radiación que emiten; por ejemplo, la acreción de materia en sistemas dobles.

Además de su interés astrofísico, los púlsares cumplen otras funciones. Una de ellas ha sido utilizarlos para comprobar la predicción de la relatividad general de que masas aceleradas emiten radiación gravitacional (un fenómeno análogo al que se produce con cargas eléctricas: la radiación electromagnética).

La confirmación de que, efectivamente, la radiación gravitacional existe derivó del descubrimiento, en 1974, del primer sistema formado por dos púlsares interactuando entre sí (denominado PSR1913+16), por el que Russell Hulse (n. 1950) y Joseph Taylor (n. 1941) recibieron en 1993 el Premio Nobel de Física. En 1978, después de varios años de observaciones continuadas de ese sistema binario, pudo concluirse que las órbitas de los púlsares

⁶ En 1974 Hewish compartió el Premio Nobel de Física con Ryle. Jocelyn Bell, que había observado por primera vez los púlsares, fue dejada al margen.

⁷ La posibilidad de que existieran estrellas de neutrones —especie de núcleos gigantes formados únicamente por neutrones y unidos por la fuerza de la gravedad— fue propuesta por primera vez en 1934 (esto es, sólo dos años después de que Chadwick descubriese el neutrón) por el físico y astrofísico suizo afincado en California (Caltech) Fritz Zwicky (1898-1974). La masa mínima para que pueda existir una estrella de neutrones es, según la relatividad general, 0,1 masas solares, mientras que el máximo parece encontrarse en torno a 6 masas solares. En el caso de una estrella de neutrones de 1 masa solar, su radio sería de unos 13 kilómetros y su densidad $2 \cdot 10^{17} \text{ kg} \cdot \text{m}^{-3}$, alrededor de $2 \cdot 10^{14}$ veces la densidad del agua.

varían acercándose entre sí, un resultado que se interpreta en términos de que el sistema pierde energía debido a la emisión de ondas gravitacionales (Taylor, Fowler y McCulloch 1979). Desde entonces han sido descubiertos otros púlsares en sistemas binarios, pero lo que aún resta es detectar la radiación gravitacional identificando su paso por instrumentos construidos e instalados en la Tierra, una empresa extremadamente difícil dado lo minúsculo de los efectos implicados: se espera que las ondas gravitacionales que lleguen a la Tierra (originadas en algún rincón del Universo en el que tenga lugar un suceso extremadamente violento) produzcan distorsiones en los detectores de no más de una parte en 10^{21} ; esto es, una pequeña fracción del tamaño de un átomo. Existen ya operativos diseñados para lograrlo: el sistema de 4 kilómetros de detectores estadounidenses denominado LIGO, por sus siglas inglesas, *Laser Interferometric Gravitational wave Observatories*.

También los cuásares resultan ser objetos muy útiles para estudiar el Universo en conjunción con la relatividad general. Alrededor de uno entre quinientos cuásares se ven implicados en un fenómeno relativista muy interesante: la desviación de la luz que emiten debido al efecto gravitacional de otras galaxias situadas entre el cuásar en cuestión y la Tierra, desde donde se observa este fenómeno, denominado «lentes gravitacionales».⁸ El efecto puede llegar a ser tan grande que se observan imágenes múltiples de un solo cuásar.

En realidad, las lentes gravitacionales no son producidas únicamente por cuásares; también lo son por grandes acumulaciones de masas (como cúmulos de galaxias) que al desviar la luz procedente de, por ejemplo, galaxias situadas tras ellas (con respecto a nosotros) dan lugar, en lugar de a una imagen más o menos puntual, a un halo de luz, a una imagen «desdoblada». Fueron observados por primera vez en 1979, cuando Walsh, Carswell y Weyman (1979) descubrieron una imagen múltiple de un cuasar en 0957+561. Posteriormente, se han tomado fotografías con el telescopio espacial Hubble de un cúmulo de galaxias situado a unos mil millones de años-luz de distancia en las que además de las galaxias que forman el cúmulo se observan numerosos arcos (trozos de aros) que se detectan con mayor dificultad debido a ser más débiles luminosamente. Estos arcos son en realidad las imágenes de galaxias mucho más alejadas de nosotros que las que constituyen el propio cúmulo, pero que observamos mediante el efecto de lente gravitacional (el cúmulo desempeña el papel de la lente que distorsiona la luz procedente de tales galaxias). Además de proporcionar nuevas evidencias en favor de la relatividad general, estas observaciones tienen el valor añadido de que la magnitud de la desviación y distorsión que se manifiesta en estos arcos luminosos es mucho mayor del que se esperaría si no hubiese nada más en el cúmulo que las galaxias que vemos en él. De hecho, las evidencias apuntan a que estos cúmulos contienen entre cinco

y diez veces más materia de la que se ve. ¿Se trata de la materia oscura de la que hablaré más adelante?

Para muchos —al menos hasta que el problema de la materia y energía oscuras pasó a un primer plano— la radiación de fondo, los púlsares y los cuásares, de los que me he ocupado en esta sección, constituyen los tres descubrimientos más importantes en la astrofísica de la segunda mitad del siglo xx. Ciertamente, lo que estos hallazgos nos dicen, especialmente en el caso de púlsares y cuásares, es que el Universo está formado por objetos mucho más sorprendentes, y sustancialmente diferentes, de los que se había supuesto existían durante la primera mitad del siglo xx. Ahora bien, cuando se habla de objetos estelares *sorprendentes* o *exóticos*, es inevitable referirse a los agujeros negros, otro de los «hijos» de la teoría de la relatividad general.

Agujeros negros

Durante décadas tras su formulación en 1915 y haber sido explotadas las predicciones de la teoría einsteiniana de la gravitación con relación al Sistema Solar (movimiento del perihelio de Mercurio, curvatura de los rayos de luz y desplazamiento gravitacional de las líneas espectrales), la relatividad general estuvo en gran medida en manos de los matemáticos, hombres como Hermann Weyl (1885-1955), Tullio Levi-Civita (1873-1941), Jan Arnouldus Schouten (1883-1971), Cornelius Lanczos (1892-1974) o André Lichnerowicz (1915-1998). La razón era, por un lado, la dificultad matemática de la teoría, y por otro el que apenas existían situaciones en las que se pudiese aplicar. Su dominio era el Universo y explorarlo requería de unos medios tecnológicos que no existían entonces (también, por supuesto, era preciso una financiación importante). Este problema fue desapareciendo a partir de finales de la década de 1960, y hoy se puede decir que la relatividad general se ha integrado plenamente en la física experimental, incluyendo apartados que nos son tan próximos como el Global Positioning System (GPS). Y no sólo en la física experimental correspondiente a los dominios astrofísico y cosmológico, también, como veremos más adelante, se ha asociado a la física de altas energías.

Y en este punto, como uno de los objetos estelares más sorprendentes y atractivos vinculados a la relatividad general cuya existencia se ha descubierto en las últimas décadas, es necesario referirse a los agujeros negros, que de hecho han ido más allá del mundo puramente científico, afincándose asimismo en el social.

Estos objetos pertenecen, como digo, al dominio teórico de la teoría de la relatividad general, aunque sus equivalentes newtonianos habían sido propuestos —y olvidados— mucho antes por el astrónomo británico John Michell (c. 1724-1793) en 1783, y por Pierre Simon Laplace (1749-1827) en 1795. Su exotividad proviene de que involucran nociones tan radicales como la destrucción del espacio-tiempo en puntos denominados «singularidades».⁹

⁸ Como en otras ocasiones, Einstein (1936) ya había predicho la existencia de este fenómeno.

⁹ Hay que recordar que desde el punto de vista de la teoría de la relatividad general espacio-tiempo y gravedad representan el mismo concepto físico, en tanto que es la curvatura del primero la que describe la segunda.

El origen de los estudios que condujeron a los agujeros negros se remonta a la década de 1930, cuando el físico de origen hindú, Subrahmanyan Chandrasekhar (1910-1995), y el ruso Lev Landau (1908-1968), mostraron que en la teoría de la gravitación newtoniana un cuerpo frío de masa superior a 1,5 veces la del Sol no podría soportar la presión producida por la gravedad (Chandrasekhar 1931; Landau 1932). Este resultado condujo a la pregunta de qué sucedería según la relatividad general. Robert Oppenheimer (1904-1967), junto a dos de sus colaboradores, George M. Volkoff y Hartland Snyder (1913-1962) demostraron en 1939 que una estrella de semejante masa se colapsaría hasta reducirse a una singularidad, esto es, a un punto de volumen cero y densidad infinita (Oppenheimer y Volkoff 1939; Oppenheimer y Snyder 1939).

Pocos prestaron atención, o creyeron, en las conclusiones de Oppenheimer y sus colaboradores y su trabajo fue ignorado hasta que el interés en los campos gravitacionales fuertes fue impulsado por el descubrimiento de los cuásares y los púlsares. Un primer paso lo dieron en 1963 los físicos soviéticos, Evgenii M. Lifshitz (1915-1985) e Isaak M. Khalatnikov (n. 1919), que comenzaron a estudiar las singularidades del espacio-tiempo relativista. Siguiendo la estela del trabajo de sus colegas soviéticos e introduciendo poderosas técnicas matemáticas, a mediados de la década de 1960 el matemático y físico británico Roger Penrose (n. 1931) y el físico Stephen Hawking (n. 1942), demostraron



Stephen Hawking.

que las singularidades eran inevitables en el colapso de una estrella si se satisfacían ciertas condiciones.¹⁰

Un par de años después de que Penrose y Hawking publicasen sus primeros artículos, la física de las singularidades del espacio-tiempo se convirtió en la de los «agujeros negros», un término afortunado que no ha hecho sino atraer la atención popular sobre este ente físico. El responsable de esta aparentemente insignificante pequeña revolución terminológica fue el físico estadounidense, John A. Wheeler (1911-2008). Él mismo explicó la génesis del término de la forma siguiente (Wheeler y Ford 1998, 296-297):

En el otoño de 1967, Vittorio Canuto, director administrativo del Instituto Goddard para Estudios Espaciales de la NASA en el 2880 de Broadway, en Nueva York, me invitó a dar una conferencia para considerar posibles interpretaciones de las nuevas y sugerentes evidencias que llegaban de Inglaterra acerca de los púlsares. ¿Qué eran estos púlsares? ¿Enanas blancas que vibraban? ¿Estrellas de neutrones en rotación? ¿Qué? En mi charla argumenté que debíamos considerar la posibilidad de que en el centro de un púlsar se encontrase un objeto completamente colapsado gravitacionalmente. Señalé que no podíamos seguir diciendo, una y otra vez, «objeto completamente colapsado gravitacionalmente». Se necesitaba una frase descriptiva más corta. ¿Qué tal agujero negro?, preguntó alguien de la audiencia. Yo había estado buscando el término adecuado durante meses, rumiándolo en la cama, en la bañera, en mi coche, siempre que tenía un momento libre. De repente, este nombre me pareció totalmente correcto. Cuando, unas pocas semanas después, el 29 de diciembre de 1967, pronuncié la más formal conferencia Sigma Xi-Phi Kappa en la West Ballroom del Hilton de Nueva York, utilicé este término, y después lo incluí en la versión escrita de la conferencia publicada en la primavera de 1968.

El nombre era sugerente y permanecería, pero la explicación era errónea (como ya he señalado un púlsar está propulsado por una estrella de neutrones).

Aunque la historia de los agujeros negros tiene sus orígenes, como se ha indicado, en los trabajos de índole física de Oppenheimer y sus colaboradores, durante algunos años predominaron los estudios puramente matemáticos, como los citados de Penrose y Hawking. La idea física subyacente era que debían representar objetos muy diferentes a cualquier otro tipo de estrella, aunque su origen estuviese ligado a éstas. Surgirían cuando, después de agotar su combustible nuclear, una estrella muy masiva comenzase a contraerse irreversiblemente debido a la fuerza gravitacional. Así, llegaría un momento en el que se formaría una región (denominada «horizonte») que únicamente dejaría entrar materia y radiación, sin permitir que saliese nada, ni siquiera luz (de ahí lo de «negro»): cuanto más grande es, más come, y cuanto más come, más crece. En el centro del agujero negro está el punto de *colapso*. De acuerdo con la relatividad general, allí la materia que una vez compuso la estrella es comprimida y expulsada aparentemente «fuera de la existencia».

Evidentemente, «fuera de la existencia» no es una idea aceptable. Ahora bien, existe una vía de escape a semejante paradójica solución: la teoría de la relatividad general no es compatible con los requisitos cuánticos,

10

Véase, por ejemplo, Penrose (1965), Hawking (1965, 1966a, 1966b) y Hawking y Penrose (1969). Aunque no me detendré en este punto, sí quiero señalar que en esta empresa también participaron otros científicos, como G. F. R. Ellis.

pero cuando la materia se comprime en una zona muy reducida son los efectos cuánticos los que dominarán. Por consiguiente, para comprender realmente la física de los agujeros negros es necesario disponer de una teoría cuántica de la gravitación (cuantizar la relatividad general o construir una nueva teoría de la interacción gravitacional que sí se pueda cuantizar), una tarea aún pendiente en la actualidad, aunque se hayan dado algunos pasos en esta dirección, uno de ellos debido al propio Hawking, el gran gurú de los agujeros negros: la denominada «radiación de Hawking» (Hawking 1975), la predicción de que, debido a procesos de índole cuántica, los agujeros negros no son tan negros como se pensaba, pudiendo emitir radiación.¹¹

No sabemos, en consecuencia, muy bien qué son estos misteriosos y atractivos objetos. De hecho, ¿existen realmente? La respuesta es que sí. Cada vez hay mayores evidencias en favor de su existencia. La primera de ellas fue consecuencia de la puesta en órbita, el 12 de diciembre de 1970, desde Kenia, para conmemorar la independencia del país, de un satélite estadounidense bautizado como Uhuru, la palabra suajili para «Libertad». Con este instrumento se pudo determinar la posición de las fuentes de rayos X más poderosas. Entre las 339 fuentes identificadas, figura Cygnus X-1, una de las más brillantes de la Vía Láctea, en la región del Cisne. Esta fuente se asoció posteriormente a una estrella supergigante azul visible de una masa 13 veces la del Sol y una compañera invisible cuya masa se estimó —analizando el movimiento de su compañera— en 7 masas solares, una magnitud demasiado grande para ser una enana blanca o una estrella de neutrones, por lo que se considera un agujero negro. No obstante, algunos sostienen que la masa de este objeto invisible es de 3 masas solares, con lo que podría ser una estrella de neutrones. En la actualidad se acepta generalmente que existen agujeros negros supermasivos en el centro de aquellas galaxias (aproximadamente el 1% del total de galaxias del Universo) cuyo núcleo es más luminoso que el resto de toda la galaxia. De manera indirecta se han determinado las masas de esos superagujeros negros en más de doscientos casos, pero sólo en unos pocos de manera directa; uno de ellos está en la propia Vía Láctea.

Inflación y «arrugas en el tiempo»

El estudio del Universo constituye un rompecabezas descomunal. Medir ahí distancias, masas y velocidades, tres datos básicos, es, obviamente, extremadamente complejo: no podemos hacerlo directamente ni tampoco podemos «ver» todo con precisión. Con los datos de que se disponía, durante un tiempo bastó con el modelo que suministraba la solución de Robertson-Walker-Friedmann de la relatividad general, que representa el Universo expandiéndose con una aceleración que depende de su contenido de masa-energía. Pero existían problemas para la cosmología del Big Bang que fueron haciéndose cada vez más patentes.

Uno de ellos era si esa masa-energía es tal que el Universo continuará expandiéndose siempre o si es lo suficien-

temente grande como para que la atracción gravitacional termine venciendo a la fuerza del estallido inicial haciendo que, a partir de un momento, comience a contraerse para finalmente llegar a un *Big Crunch* (Gran Contracción). Otro problema residía en la gran uniformidad que se observa en la distribución de masa del Universo si uno toma como unidad de medida escalas de unos 300 millones de años-luz o más (a pequeña escala, por supuesto, el Universo, con sus estrellas, galaxias, cúmulos de galaxias y enormes vacíos interestelares, no es homogéneo). El fondo de radiación de microondas es buena prueba de esa macro-homogeneidad. Ahora bien, en la teoría estándar del Big Bang es difícil explicar esta homogeneidad mediante los fenómenos físicos conocidos; además, si tenemos en cuenta que la transmisión de información sobre lo que sucede entre diferentes puntos del espacio-tiempo no puede ser transmitida con una velocidad superior a la de la luz, sucede que en los primeros momentos de existencia del Universo no habría sido posible que regiones distintas «llegasen a un consenso», por decirlo de alguna manera, acerca de cuál debería ser la densidad media de materia y radiación.¹²

Para resolver este problema se propuso la idea del Universo inflacionario, según la cual en los primeros instantes de vida del Universo se produjo un aumento gigantesco, exponencial, en la velocidad de su expansión. En otras palabras, el miniUniverso habría experimentado un crecimiento tan rápido que no habría habido tiempo para que se desarrollasen procesos físicos que diesen lugar a distribuciones inhomogéneas. Una vez terminada la etapa inflacionaria, el Universo habría continuado evolucionando de acuerdo con el modelo clásico del Big Bang.

En cuanto a quiénes fueron los científicos responsables de la teoría inflacionaria, los principales nombres que hay que citar son los del estadounidense Alan Guth (n. 1947) y el soviético Andrei Linde (n. 1948).¹³ Pero más que nombres concretos, lo que me interesa resaltar es que no es posible comprender esta teoría al margen de la física de altas energías (antes denominada de partículas elementales), de la que me ocuparé más adelante; en concreto de las denominadas teorías de gran unificación (Grand Unified Theories; GUT), que predicen que tendría que producirse una transición de fase a temperaturas del orden de 10^{27} grados Kelvin.¹⁴ Aquí tenemos una muestra de uno de los fenómenos más importantes que han tenido lugar en la física de la segunda mitad del siglo xx: la reunión de la cosmología, la ciencia de «lo grande», y la física de altas energías, la ciencia de «lo pequeño»; naturalmente, el lugar de encuentro ha sido los primeros instantes de vida del Universo, cuando las energías implicadas fueron gigantescas.

Bien, la inflación da origen a un Universo uniforme, pero entonces ¿cómo surgieron las minúsculas inhomogeneidades primordiales de las que habrían nacido, al pasar el tiempo y actuar la fuerza gravitacional, estructuras cósmicas como las galaxias?

11

Semejante emisión conduciría a una lenta disminución de la masa del agujero negro. Si la disminución fuese continua, el agujero negro podría terminar desapareciendo. Sin embargo, para agujeros negros normales (de unas pocas masas solares), esto no sucedería. En el caso, por ejemplo, de un agujero negro de una masa solar, su temperatura sería menor que la de la radiación del fondo de microondas, lo que significa que agujeros negros de semejante masa absorberían radiación más rápidamente de la que emitirían; esto es, continuarían aumentando de masa. Sin embargo, si existiesen agujeros negros muy pequeños (formados, por ejemplo, en los primeros instantes del Universo, debido a las fluctuaciones de densidad que se debieron producir entonces), éstos tendrían una temperatura mucho mayor y radiarían más energía de la que absorberían. Perderían masa, lo que les haría todavía más calientes, estallando finalmente en una gran explosión de energía. Su vida sería tal que quizás podríamos observar tales explosiones ahora. No obstante, aún no han sido detectadas.

12

Esta dificultad recibe el nombre de «problema del horizonte».

13

Guth (1981), Linde (1982).

14

En una transición de fase tiene lugar un cambio repentino en el estado del sistema en cuestión. Un ejemplo es cuando el agua (líquida) se convierte en hielo (sólido).

Una posible respuesta a esta pregunta era que la inflación podría haber amplificado enormemente las ultramicroscópicas fluctuaciones cuánticas que se producen debido al principio de incertidumbre aplicado a energías y tiempo ($\Delta E \Delta t \geq \hbar$). Si era así, ¿dónde buscar tales inhomogeneidades mejor que en el fondo de radiación de microondas?

La respuesta a esta cuestión vino de los trabajos de un equipo de científicos estadounidenses a cuya cabeza estaban John C. Mather (n. 1946) y George Smoot (n. 1945). Cuando la NASA aprobó en 1982 fondos para la construcción de un satélite –el Cosmic Background Explorer (COBE), que fue puesto en órbita, a 900 kilómetros de altura, en el otoño de 1989– para estudiar el fondo cósmico de microondas, Mather se encargó de coordinar todo el proceso, así como del experimento (en el que utilizó un espectrofotómetro enfriado a 1,5°K) que demostró que la forma del fondo de radiación de microondas se ajustaba a la de una radiación de cuerpo negro a la temperatura de 2,735°K, mientras que Smoot midió las minúsculas irregularidades predichas por la teoría de la inflación. Diez años después, tras haber intervenido en los trabajos más de mil personas y con un coste de 160 millones de dólares, se anunciaba (Mather et al. 1990; Smoot et al. 1992) que el COBE había detectado lo que Smoot denominó «arrugas» del espacio-tiempo, las semillas de las que surgieron las complejas estructuras –como las galaxias– que ahora vemos en el Universo.¹⁵ Podemos captar algo de la emoción que sintieron estos investigadores al comprobar sus resultados a través de un libro de divulgación que Smoot publicó poco después, *Wrinkles in Time (Arrugas en el tiempo)*. Escribió allí (Smoot y Davidson 1994, 336):

Estaba contemplando la forma primordial de las arrugas, podía sentirlo en mis huesos. Algunas de las estructuras eran tan grandes que sólo podían haber sido generadas durante el nacimiento del Universo, no más tarde. Lo que tenía ante mí era la marca de la creación, las semillas del Universo presente.

En consecuencia, «la teoría del Big Bang era correcta y la de la inflación funcionaba; el modelo de las arrugas encajaba con la formación de estructuras a partir de la materia oscura fría; y la magnitud de la distribución habría producido las estructuras mayores del Universo actual bajo el influjo del colapso gravitacional a lo largo de 15.000 millones de años».

El COBE fue un magnífico instrumento, pero en modo alguno el único. Los ejemplos en los que astrofísica y tecnología se dan la mano son múltiples. Y no sólo instrumentos instalados en la Tierra, también vehículos espaciales. Así, hace ya bastante que el Sistema Solar se ve frecuentado por satélites con refinados instrumentos que nos envían todo tipo de datos e imágenes. Sondas espaciales como Mariner 10, que observó, en 1973, Venus desde 10.000 kilómetros; Pioneer 10 y Voyager 1 y 2, que entre 1972 y 1977 se adentraron por los alrededores de Júpiter, Saturno, Urano y Plutón, o Galileo, dirigido hacia Júpiter y sus satélites.

Un tipo muy especial de vehículo es el telescopio espacial Hubble, que la NASA puso en órbita, después de un largo proceso, en la primavera de 1990.¹⁶ Situar un telescopio en un satélite artificial significa salvar ese gran obstáculo para recibir radiaciones que es la atmósfera terrestre. Desde su lanzamiento, y especialmente una vez que se corrigieran sus defectos, el Hubble ha enviado y continúa enviando imágenes espectaculares del Universo. Gracias a él, por primera vez disponemos de fotografías de regiones (como la nebulosa de Orión) en las que parece que se está formando una estrella. No es completamente exagerado decir que ha revolucionado nuestro conocimiento del Universo.

Planetas extrasolares

Gracias al avance tecnológico los científicos están siendo capaces de ver nuevos aspectos y objetos del Cosmos, como, por ejemplo, la existencia de sistemas planetarios asociados a estrellas que no sean el Sol. El primer hallazgo en este sentido se produjo en 1992, cuando Alex Wolszczan y Dale Frail descubrieron que al menos dos planetas del tipo de la Tierra orbitan alrededor de un púlsar (Wolszczan y Frail 1992); tres años después, Michel Mayor y Didier Queloz hicieron público que habían descubierto un planeta del tamaño y tipo de Júpiter (un gigante gaseoso) orbitando en torno a la estrella 51 Pegasi (Mayor y Queloz 1995). Desde entonces el número de planetas extrasolares conocidos ha aumentado considerablemente. Y si existen tales planetas, acaso en algunos también se haya desarrollado vida. Ahora bien, aunque la biología que se ocupa del problema del origen de la vida no descarta que en entornos lo suficientemente favorables las combinaciones de elementos químicos puedan producir, debido a procesos sinérgicos, vida, ésta no tiene porque ser vida del tipo de la humana. La biología evolucionista, apoyada en los registros geológicos, ha mostrado que la especie humana es producto del azar evolutivo. Si, por ejemplo, hace 65 millones de años no hubiese chocado contra la Tierra, a una velocidad de aproximadamente treinta kilómetros por segundo, un asteroide o un cometa de unos diez kilómetros de diámetro, produciendo una energía equivalente a la que librarían cien millones de bombas de hidrógeno, entonces acaso no habrían desaparecido (no, desde luego, entonces) una cantidad enorme de especies vegetales y animales, entre las que se encontraban los dinosaurios, que no dejaban prosperar a los, entonces, pequeños mamíferos, que con el paso del tiempo terminarían produciendo, mediante procesos evolutivos, especies como la de los *homo sapiens*.

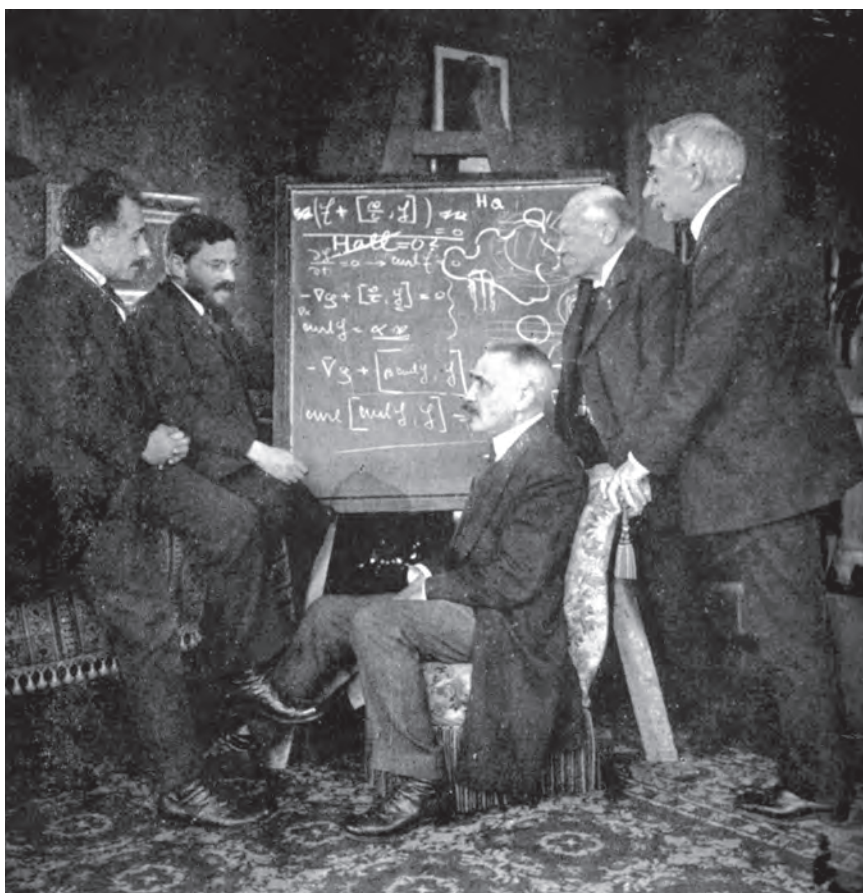
Precisamente por semejante aleatoriedad es por lo que no podemos estar seguros de que exista en otros planetas, en nuestra o en otra galaxia, vida inteligente que trate, o haya tratado, de entender la naturaleza construyendo sistemas científicos, y que también se haya planteado el deseo de comunicarse con otros seres vivos que puedan existir

15

Por sus trabajos ambos recibieron el Premio Nobel de Física de 2006.

16

Hasta comienzos de la década de 1990 el tamaño de los espejos de los telescopios ópticos instalados en la Tierra más grandes era de entre 5 y 6 metros de diámetro. El mayor telescopio de entonces, con un espejo primario (o colector) de 6 metros de diámetro, estaba en el Cáucaso ruso. Venía después el telescopio del observatorio de Monte Palomar, inaugurado poco después de la Segunda Guerra Mundial, con un diámetro de 5 metros, y a continuación una larga serie de telescopios de en torno a 4 metros. En la actualidad se han completado o están en marcha una serie de proyectos de construcción de grandes telescopios que, por sus características y su aprovechamiento de las tecnologías más modernas, están permitiendo ya dar un importante salto cuantitativo –y en más de un sentido cualitativo también– en la investigación astrofísica. Se trata de telescopios de hasta 10 metros, como el que ya está en funcionamiento en Manua Kea, Hawai, propiedad del California Institute of Technology y de la Universidad de California, y otro que se está construyendo en el mismo lugar. Con el construido, el más grande del mundo, se observó, en el cúmulo de las Pléyades, una enana marrón (PPL15), una estrella tan pequeña que no luce como las demás, por lo que hasta entonces no habían sido vistas, aunque sí detectadas en algún caso a través de sus efectos gravitacionales. Otro instrumento de este tamaño es el Gran Telescopio del Instituto de Astrofísica de Canarias, instalado en el Roque de los Muchachos, que ya ha visto la primera luz.



Albert Einstein junto a Paul Ehrenfest, Paul Langevin, Heike Kamerlingh Onnes y Pierre Weiss en la casa de Ehrenfest (Leiden, 1920).

en el Universo. Aun así, desde hace tiempo existen programas de investigación que rastrean el Universo buscando señales de vida inteligente. Programas como el denominado SETI, siglas del Search of Extra-Terrestrial Intelligence (Búsqueda de Inteligencia Extraterrestre), que ha utilizado receptores con 250 millones de canales, que realizan alrededor de veinte mil millones de operaciones por segundo.

Materia y energía oscuras

La existencia de planetas extrasolares ciertamente nos conmueve y emociona, pero no es algo «fundamental»; no altera los pilares del edificio científico. Muy diferente es el caso de otros descubrimientos relativos a los contenidos del Universo. Me estoy refiriendo a que tenemos buenas razones para pensar que existe en el Cosmos una gran cantidad de materia que no observamos, pero que ejerce fuerza gravitacional. La evidencia más inmediata procede de galaxias en forma de disco (como nuestra propia Vía Láctea) que se encuentran en rotación. Si miramos a la parte exterior de estas galaxias, encontramos que el gas se mueve de manera sorprendentemente rápida; mucho más rápidamente de lo que debería debido a la atracción gravitacional producida por las estrellas y gases que detectamos en su interior. Otras evidencias proceden de los movimientos internos de cúmulos de galaxias. Se cree que

esta materia «oscura» constituye el 30% de toda la materia del Universo. ¿Cuál es su naturaleza? Ése es uno de los problemas; puede tratarse de estrellas muy poco luminosas (como las enanas marrones), de partículas elementales exóticas o de agujeros negros. No podremos entender realmente lo que son las galaxias, o cómo se formaron, hasta que sepamos qué es esa materia oscura. Ni tampoco podremos saber cuál será el destino último de nuestro Universo.

Junto al problema de la materia oscura, otro parecido ha adquirido prominencia en la última década del siglo xx: el de la energía oscura. Estudiando un tipo de supernovas –estrellas que han explotado dejando un núcleo–, un grupo dirigido por Saul Perlmutter (del Laboratorio Lawrence en Berkeley, California) y otro por Brian Schmidt (Observatorios de Monte Stromlo y Siding Spring, en Australia) llegaron a la conclusión de que, al contrario de lo supuesto hasta entonces, la expansión del Universo se está acelerando (Perlmutter et al. 1998; Schmidt et al. 1998). El problema es que la masa del Universo no puede explicar tal aceleración; había que suponer que la gravedad actuaba de una nueva y sorprendente manera: alejando las masas entre sí, no atrayéndolas. Se había supuesto que para propulsar el Big Bang debía de haber existido una energía repulsiva en la creación del Universo, pero no se había pensado que pudiera existir en el Universo ya maduro.

Una nueva energía entraba así en acción, una energía «oscura» que reside en el espacio vacío. Y como la energía es equivalente a la masa, esta energía oscura significa una nueva aportación a la masa total del Universo, distinta, eso sí, de la masa oscura. Se tiene, así, que alrededor del 30% del Universo está formado por masa ordinaria, el 30% de masa oscura y el 67% de energía oscura. En otras palabras: creíamos que conocíamos eso que llamamos Universo y resulta que es un gran desconocido. Porque ni sabemos qué es la materia oscura ni lo que es la energía oscura. Una posible explicación de esta última se podría encontrar en un término que introdujo Einstein en 1916-1917 en las ecuaciones de campo de la relatividad general. Como vimos, al aplicar su teoría de la interacción gravitacional al conjunto del Universo, Einstein buscaba encontrar un modelo que representase un Universo estático y ello le obligó a introducir en sus ecuaciones un nuevo término, la ya citada constante cosmológica, que en realidad representaba un campo de fuerza repulsiva, para compensar el efecto atractivo de la gravitación. Al encontrarse soluciones de la cosmología relativista que representan un Universo en expansión y demostrarse observacionalmente (Hubble) que el Universo se expande, Einstein pensó que no era necesario mantener aquella constante, aunque podía incorporarse sin ningún problema en los modelos expansivos teóricos. Acaso ahora sea necesario resucitarla. Ahora bien, semejante resurrección no se podrá limitar a incluirla de nuevo en la cosmología relativista; esto ya no basta: es preciso que tome su sentido y lugar en las teorías cuánticas que intentan insertar

la gravitación en el edificio cuántico; al fin y al cabo la energía oscura es la energía del vacío, y éste tiene estructura desde el punto de vista de la física cuántica. Y puesto que ha salido, una vez más, la física cuántica es hora de pasar a ella, a cómo se desarrolló y consolidó la revolución cuántica durante la segunda mitad del siglo xx.

Un mundo cuántico

La física de altas energías: de los protones, neutrones y electrones a los quarks

Antes, al tratar de la revolución cuántica que surgió en la primera mitad del siglo, me referí a la búsqueda de los componentes básicos de la materia, las denominadas «partículas elementales». Vimos entonces cómo ir más allá de protones, electrones y neutrones, las más básicas de esas partículas, requería energías más elevadas de las que podían proporcionar los «proyectiles» –por ejemplo, partículas alfa– que proporcionaban las emisiones de elementos radiactivos (especialmente el radio), y que fue Ernest Lawrence quien abrió una nueva senda introduciendo y desarrollando unos instrumentos denominados aceleradores de partículas (ciclotrones en su caso), cuyo funcionamiento se basa en acelerar partículas a energías elevadas, haciéndolas chocar luego unas con otras (o con algún blanco predeterminado) para ver qué es lo que se produce en tales choques; esto es, de qué nuevos componentes más pequeños están compuestas esas partículas... si es que lo están.¹⁷

La física de partículas elementales, también llamada, como ya indiqué, de altas energías, ha sido una de las grandes protagonistas de la segunda mitad del siglo xx. Se trata de una ciencia muy cara (es el ejemplo canónico de *Big Science*, Gran Ciencia, ciencia que requiere de grandes equipos de científicos y técnicos y de grandes inversiones); cada vez, de hecho, más cara, al ir aumentando el tamaño de los aceleradores para poder alcanzar mayores energías. Después de la Segunda Guerra Mundial contó –especialmente en Estados Unidos– con la ayuda del prestigio de la física nuclear, que había suministrado las poderosas bombas atómicas. Limitándome a citar los aceleradores más importantes construidos, señalaré que en 1952 entró en funcionamiento en Brookhaven (Nueva York) el denominado Cosmotrón, para protones, que podía alcanzar 2,8 GeV;¹⁸ luego vinieron, entre otros, el Bevatrón (Berkeley, protones; 1954), 3,5 GeV; Dubna (URSS, protones; 1957), 4,5 GeV; Proton-Synchrotron (CERN, Ginebra, protones; 1959), 7 GeV; SLAC (Stanford; 1966), 20 GeV; PETRA (Hamburgo, electrones y positrones; 1978), 38 GeV; Collider (CERN, protones y antiprotones; 1981), 40 GeV; Tevatron (Fermilab, Chicago, protones y antiprotones), 2.000 GeV, y SLC (Stanford, California, electrones y positrones), 100 GeV, los dos de 1986; LEP (CERN, electrones y positrones; 1987), 100 GeV, y HERA (Hamburgo, electrones y protones; 1992), 310 GeV.

Las siglas CERN corresponden al Centre Européen de Recherches Nucleaires (Centro Europeo de Investigaciones Nucleares), la institución que en 1954 crearon en Ginebra doce naciones europeas para poder competir con Estados Unidos. Con sus aceleradores, el CERN –formado ahora por un número mayor de países (España es uno de ellos)– ha participado de manera destacada en el desarrollo de la física de altas energías. De hecho, en tiempos difíciles para esta disciplina como son los actuales, el CERN acaba de completar (2008) la construcción de un nuevo acelerador, uno en el que los protones chocarán con una energía de 14.000 GeV: el LHC (Large Hadron Collider). Toma así la vieja Europa la antorcha en mantener el «fuego» de esta costosa rama de la física.

¿Por qué he dicho «tiempos difíciles para esta disciplina»? Pues porque debido a su elevado costo, en los últimos años esta rama de la física está pasando por dificultades. De hecho, hace poco sufrió un duro golpe en la que hasta entonces era su patria principal, Estados Unidos. Me estoy refiriendo al Supercolisionador Superconductor (Superconducting Super Collider; SSC). Este gigantesco acelerador, que los físicos de altas energías norteamericanos estimaban indispensable para continuar desarrollando la estructura del denominado modelo estándar, iba a estar formado por un túnel de 84 kilómetros de longitud que debería ser excavado en las proximidades de una pequeña población de 18.000 habitantes, situada a 30 kilómetros al sudoeste de Dallas, en Waxahachie. En el interior de ese túnel miles de bobinas magnéticas superconductoras guiarían dos haces de protones para que, después de millones de vueltas, alcanzarán una energía veinte veces más alta que la que se podía conseguir en los aceleradores existentes. En varios puntos a lo largo del anillo, los protones de los dos haces chocarían, y enormes detectores controlarían lo que sucediera en tales colisiones. El coste del proyecto, que llevaría diez años, se estimó inicialmente en 6.000 millones de dólares.

Después de una azarosa vida, con parte del trabajo de infraestructura ya realizado (la excavación del túnel), el 19 de octubre de 1993 y después de una prolongada, difícil y cambiante discusión parlamentaria, tanto en el Congreso como en el Senado, el Congreso canceló el proyecto. Otros programas científicos –especialmente en el campo de las ciencias biomédicas– atraían la atención de los congresistas y senadores americanos; y también, ¿cómo negarlo?, de la sociedad, más interesada en asuntos que atañen a su salud.

Pero dejemos los aceleradores y vayamos a su producto, a las partículas aparentemente «elementales». Gracias a los aceleradores, su número fue creciendo de tal manera que terminó socavando drásticamente la idea de que la mayoría pudiesen ser realmente elementales en un sentido fundamental. Entre las partículas halladas podemos recordar, por ejemplo, piones y muones de diversos tipos, o las denominadas Λ , W o Z , sin olvidar a sus correspondientes antipartículas.¹⁹ El número –cientos– resultó ser tan eleva-

17

Estrictamente no fue Lawrence quien inició el camino de la física de partículas elementales utilizando fuentes no radiactivas, aunque sí quien dio con el procedimiento técnico más adecuado. Antes, en Cambridge, John D. Cockcroft (1897-1967) y Ernest T. S. Walton (1903-1995) utilizaron un multiplicador voltaico que les proporcionó los 500 kV (1 kV=1.000 voltios) que necesitaron para ser los primeros en observar, en 1932, la desintegración artificial de átomos de litio en dos partículas α . Y hubo más precedentes, como los generadores desarrollados por Robert J. Van de Graaff (1901-1967).

18

1 GeV=1.000 millones de electrón-voltios. 1 electrón-voltio es la energía de movimiento que ganaría un electrón sometido a la diferencia de potencial de un voltio.

19

Cada partícula tiene su antipartícula (aunque a veces coincidan): cuando una y otra se encuentran, desaparecen –se aniquilan– produciendo energía.

do que llegó a hablarse de un «zoo de partículas», un zoo con una fauna demasiado elevada.

A ese zoo se les unió otra partícula particularmente llamativa: los quarks. Su existencia fue propuesta teóricamente en 1964 por los físicos estadounidenses Murray Gell-Mann (n. 1929) y George Zweig (n. 1937). Hasta su aparición en el complejo y variado mundo de las partículas elementales, se pensaba que protones y neutrones eran estructuras atómicas inquebrantables, realmente básicas, y que la carga eléctrica asociada a protones y electrones era una unidad indivisible. Los quarks no obedecían a esta regla, ya que se les asignó cargas fraccionarias. De acuerdo a Gell-Mann (1964) y Zweig (1964), los hadrones, las partículas sujetas a la interacción fuerte, están formados por dos o tres especies de quarks y antiquarks, denominados *u* (*up*; arriba), *d* (*down*; abajo) y *s* (*strange*; extraño), con, respectivamente, cargas eléctricas $2/3$, $-1/3$ y $-1/3$ la del electrón.²⁰ Así, un protón está formado por dos quarks *u* y uno *d*, mientras que un neutrón está formado por dos quarks *d* y por otro *u*; son, por consiguiente, estructuras compuestas. Posteriormente, otros físicos propusieron la existencia de tres quarks más: *charm* (*c*; 1974), *bottom* (*b*; 1977) y *top* (*t*; 1995). Para caracterizar esta variedad, se dice que los quarks tienen seis tipos de «sabores» (*flavours*); además, cada uno de estos seis tipos puede ser de tres clases, o *colores*: rojo, amarillo (o verde) y azul. Y para cada quark existe, claro, un antiquark.

Por supuesto, nombres como los anteriores –color, sabor, arriba, abajo...– no representan la realidad que asociamos normalmente a tales conceptos, aunque puede en algún caso existir una cierta lógica en ellos, como sucede con el *color*. Veamos lo que el propio Gell-Mann (1995, 199) ha señalado con respecto a este término:

Aunque el término «color» es más que nada un nombre gracioso, sirve también como metáfora. Hay tres colores, denominados rojo, verde y azul a semejanza de los tres colores básicos en una teoría simple de la visión humana del color (en el caso de la pintura, los tres colores primarios suelen ser el rojo, el amarillo y el azul, pero para mezclar luces en lugar de pigmentos, el amarillo se sustituye por el verde). La receta para un neutrón o un protón consiste en tomar un quark de cada color, es decir, uno rojo, uno verde y uno azul, de modo que la suma de colores se anule. Como en la visión del color blanco se puede considerar una mezcla de rojo, verde y azul, podemos decir metafóricamente que el neutrón y el protón son blancos.

En definitiva, los quarks tienen *color* pero los hadrones no: son *blancos*. La idea es que sólo las partículas *blancas* son observables directamente en la naturaleza, mientras que los quarks no; ellos están «confinados», asociados formando hadrones. Nunca podremos observar un quark libre. Ahora bien, para que los quarks permanezcan confinados deben existir fuerzas entre ellos muy diferentes de las electromagnéticas o de las restantes. «Así como la fuerza electromagnética entre electrones está mediada por el intercambio virtual de fotones», utilizando de nuevo a Gell-Mann (1995, 200), «los quarks están ligados entre sí por una fuerza que surge del intercambio de otros cuantos: los

gluones (del inglés *glue*, pegar), llamados así porque hacen que los quarks se peguen formando objetos observables blancos como el protón y el neutrón».²¹

Aproximadamente una década después de la introducción de los quarks, se desarrolló una teoría, la cromodinámica cuántica, que explica por qué los quarks están confinados tan fuertemente que nunca pueden escapar de las estructuras hadrónicas que forman. Por supuesto, el nombre *cromodinámica* –procedente del término griego *chromos* (color)– aludía al *color* de los quarks (y el adjetivo «cuántica» a que es compatible con los requisitos cuánticos). Al ser la cromodinámica cuántica una teoría de las partículas elementales con color, y al estar éste asociado a los quarks, que a su vez tratan de los hadrones, las partículas sujetas a la interacción fuerte, tenemos que la cromodinámica cuántica describe esta interacción.

Con la electrodinámica cuántica –logro, como ya indiqué, de la primera mitad del siglo xx– y la cromodinámica cuántica, se disponía de teorías cuánticas para las interacciones electromagnética y fuerte. Pero ¿y la débil, la responsable de los fenómenos radiactivos? En 1932, Enrico Fermi (1901–1954), uno de los mejores físicos de su siglo, desarrolló una teoría para la interacción débil (que aplicó, sobre todo, a la denominada «desintegración beta», proceso radiactivo en el que un neutrón se desintegra dando lugar a un protón, un electrón y un antineutrino), que mejoraron en 1959 Robert Marshak (1916–1992), E. C. George Sudarshan (n. 1931), Richard Feynman y Murray Gell-Mann, pero la versión más satisfactoria para una teoría cuántica de la interacción débil llegó cuando en 1967 el estadounidense Steven Weinberg (n. 1933) y el año siguiente el paquistaní (afincando en Inglaterra) Abdus Salam (1929–1996) propusieron independientemente una teoría que unificaba las interacciones electromagnética y débil. Su modelo incorporaba ideas propuestas en 1960 por Sheldon Glashow (n. 1932).²² Por estos trabajos, Weinberg, Salam y Glashow compartieron el Premio Nobel de Física de 1979; esto es, después de que, en 1973, una de las predicciones de su teoría –la existencia de las denominadas «corrientes neutras débiles»– fuese corroborada experimentalmente en el CERN.

La teoría electrodébil unificaba la descripción de las interacciones electromagnética y débil, pero ¿no sería posible avanzar por la senda de la unificación, encontrando una formulación que incluyese también a la interacción fuerte, descrita por la cromodinámica cuántica? La respuesta, positiva, a esta cuestión vino de la mano de Howard Georgi (n. 1947) y Glashow, que introdujeron las primeras ideas de lo que se vino a denominar teorías de gran unificación (GUT), que ya mencioné con anterioridad (Georgi y Glashow 1974).

El impacto principal de esta familia de teorías se ha producido en la cosmología; en concreto en la descripción de los primeros instantes del Universo. Desde la perspectiva de las GUTs, al principio existía sólo una fuerza que englobaba

20

Los hadrones pueden ser de dos tipos: bariones (protones, neutrones e hiperones) y mesones (partículas cuyas masas tienen valores entre la del electrón y la del protón).

21

Es interesante también citar lo que Gell-Mann (1995, 198) ha señalado a propósito del nombre quark: «En 1963, cuando bauticé con el nombre de "quark" a los constituyentes elementales de los nucleones, partí de un sonido que no se escribía de esa forma, algo parecido a "cuorc". Entonces, en una de mis lecturas ocasionales de *Finnegans Wake*, de James Joyce, descubrí la palabra "quark" en la frase "Tres quarks para Muster Mark". Dado que "quark" (que se aplica más que nada al grito de una gaviota) estaba para rimar con "Mark", tenía que buscar alguna excusa para pronunciarlo como "cuorc". Pero el libro narra los sueños de un tabernero llamado Humphrey Chipden Earkwicker. Las palabras del texto suelen proceder simultáneamente de varias fuentes, como las "palabras híbridas" en *A través del espejo*, de Lewis Carroll. De vez en cuando aparecen frases parcialmente determinadas por la jerga de los bares. Razoné, por tanto, que tal vez una de las fuentes de la expresión "Tres quarks para Muster Mark" podría ser "Tres cuartos para Mister Mark" (cuarto en inglés es *quart*) en cuyo caso la pronunciación "cuorc" no estaría totalmente injustificada. En cualquier caso, el número tres encaja perfectamente con el número de quarks presentes en la naturaleza».

22

Glashow (1960), Weinberg (1967), Salam (1968).

las electromagnética, débil y fuerte, que fueron separándose al irse enfriando el Universo. Con semejante equipaje teórico es posible ofrecer explicaciones de cuestiones como el hecho de que exista (al menos aparentemente y para nuestra fortuna) más materia que antimateria en el Universo. Esto es debido a que las GUTs tienen en común que en ellas no se conserva una magnitud denominada «número bariónico», lo que significa que son posibles procesos en los que el número de bariones –entre los que se encuentran, recordemos, los protones y los neutrones– producidos no es igual al de antibariones. Utilizando esta propiedad, el físico japonés Motohiko Yoshimura (1978) demostró que un estado inicial en el que exista igual número de materia y antimateria puede evolucionar convirtiéndose en uno con más protones o neutrones que sus respectivas antipartículas, produciendo así un Universo como el nuestro, en el que hay más materia que antimateria.

Gracias al conjunto formado por las anteriores teorías, poseemos un marco teórico extraordinario para entender de qué está formada la naturaleza. Su capacidad predictiva es increíble. De acuerdo con él, se acepta que toda la materia del Universo está formada por agregados de tres tipos de partículas elementales: electrones y sus parientes (las partículas denominadas muón y tau), neutrinos (neutrino electrónico, muónico y tauónico) y quarks, además de por los cuantos asociados a los campos de las cuatro fuerzas que reconocemos en la naturaleza:²³ el fotón para la interacción electromagnética, las partículas Z y W (bosones gauge) para la débil, los gluones para la fuerte y, aunque la gravitación todavía no se ha incorporado a ese marco, los aún no observados gravitones para la gravitacional. El subconjunto formado por la cromodinámica cuántica y teoría electrodébil (esto es, el sistema teórico que incorpora las teorías relativistas y cuánticas de las interacciones fuerte, electromagnética y débil) es especialmente poderoso si tenemos en cuenta el balance predicciones-comprobaciones experimentales. Es denominado *modelo estándar*. De acuerdo al distinguido físico e historiador de la ciencia, Silvan Schweber (1997, 645), «la formulación del Modelo Estándar es uno de los grandes logros del intelecto humano, uno que rivaliza con la mecánica cuántica. Será recordado –junto a la relatividad general, la mecánica cuántica y el desciframiento del código genético– como uno de los avances intelectuales más sobresalientes del siglo xx. Pero, mucho más que la relatividad general y la mecánica cuántica, es el producto de un esfuerzo colectivo». Quiero hacer hincapié en esta última expresión, «esfuerzo colectivo». El lector atento de estas páginas se dará cuenta fácilmente, sin embargo, de que yo únicamente me he referido en estas páginas a unos pocos físicos; a la punta de un gran iceberg. Ha sido inevitable: la historia de la física de altas energías requiere no ya de un extenso libro, sino de varios.

Ahora bien, no obstante su éxito obviamente el modelo estándar no es «la teoría final». Por una parte porque

la interacción gravitacional queda al margen, pero también porque incluye demasiados parámetros que hay que determinar experimentalmente. Se trata de las, siempre incómodas pero fundamentales, preguntas del tipo «¿por qué?». ¿Por qué existen las partículas fundamentales que detectamos? ¿Por qué esas partículas tienen las masas que tienen? ¿Por qué, por ejemplo, el tau pesa alrededor de 3.520 veces lo que el electrón? ¿Por qué son cuatro las interacciones fundamentales, y no tres, cinco o sólo una? ¿Y por qué tienen estas interacciones las propiedades (como intensidad o rango de acción) que poseen?

¿Un mundo de ultraminúsculas cuerdas?

Pasando ahora a la gravitación, la otra interacción básica, ¿no se puede unificar con las otras tres? Un problema central es la inexistencia de una teoría cuántica de la gravitación que haya sido sometida a pruebas experimentales. Existen candidatos para cumplir ese espléndido sueño unificador, unos complejos edificios matemáticos llamados teorías de cuerdas.

Según la teoría de cuerdas, las partículas básicas que existen en la naturaleza son en realidad filamentos unidimensionales (cuerdas extremadamente delgadas) en espacios de muchas más dimensiones que las tres espaciales y una temporal de las que somos conscientes; aunque más que decir «son» o «están constituidas» por tales cuerdas, habría que decir que «son manifestaciones» de vibraciones de esas cuerdas. En otras palabras, si nuestros instrumentos fuesen suficientemente poderosos, lo que veríamos no serían «puntos» con ciertas características a los que llamamos electrón, quark, fotón o neutrino, por ejemplo, sino minúsculas cuerdas (cuyos cabos pueden estar abiertos o cerrados) vibrando. La imagen que suscita esta nueva visión de la materia más que «física» es, por consiguiente, «musical»: «Del mismo modo que las diferentes pautas vibratorias de la cuerda de un violín dan lugar a diferentes notas musicales, *los diferentes modelos vibratorios de una cuerda fundamental dan lugar a diferentes masas y cargas de fuerzas...* El Universo –que está compuesto por un número enorme de esas cuerdas vibrantes–. Es algo semejante a una sinfonía cósmica», ha señalado el físico, y miembro destacado de la «comunidad de cuerdas», Brian Greene (2001, 166-168) en un libro titulado *El Universo elegante*, que ha sido un éxito editorial.

Es fácil comprender el atractivo que algunos sienten por estas ideas: «Las cuerdas son verdaderamente fundamentales; son "átomos", es decir *componentes indivisibles*, en el sentido más auténtico de la palabra griega, tal como la utilizaron los antiguos griegos. Como componentes absolutamente mínimos de cualquier cosa, representan el final de la línea –la última de las muñecas rusas llamadas "matrioskas"– en las numerosas capas de subestructuras dentro del mundo microscópico» (Green 2001, 163). Ahora bien, ¿qué tipo de materialidad es la de estos construc-

23

Para entenderla la idea de cuanto de una interacción basta con pensar en el ya citado caso de la radiación electromagnética, que según la teoría clásica se propaga a través de campos (ondas), mientras que de acuerdo con la física cuántica lo hace mediante los corpúsculos (fotones), *cuantos* de energía $h\nu$ propuestos en 1905 por Einstein.

tos teóricos unidimensionales? ¿Podemos pensar en ellos como una especie de «materia elemental» en algún sentido parecido a aquel en el que se piensa cuando se habla habitualmente de materia, incluso de partículas tan (a la postre acaso sólo aparentemente) elementales como un electrón, un muón o un quark?

Decía antes que las teorías de cuerdas son unos complejos edificios matemáticos, y así es. De hecho, las matemáticas de la teoría de las cuerdas son tan complicadas que hasta ahora nadie conoce ni siquiera las ecuaciones de las fórmulas exactas de esa teoría, únicamente unas aproximaciones de dichas ecuaciones, e incluso estas ecuaciones aproximadas resultan ser tan complicadas que hasta la fecha sólo se han resuelto parcialmente. No es por ello sorprendente que uno de los grandes líderes de esta disciplina sea un físico especialmente dotado para las matemáticas. Me estoy refiriendo al estadounidense Edward Witten (n. 1951). Para hacerse una idea de su talla como matemático basta con señalar que en 1990 recibió (junto a Pierre-Louis Lions, Jean-Christophe Yoccoz y Shigefumi Mori) una de las cuatro medallas Fields que se conceden cada cuatro años y que constituyen el equivalente en matemáticas de los Premios Nobel. Fue Witten (1995) quien argumentó, iniciando así lo que se denomina «la segunda revolución de la cuerdas», que para que la teoría de cuerdas (o supercuerdas) pueda aspirar a ser realmente una Teoría del Todo, debe tener diez dimensiones espaciales más una temporal, esto es, once (Witten denominó Teoría M a esa formulación, todavía por desarrollar completamente).²⁴

Enfrentados con las teorías de cuerdas, es razonable preguntarse si al avanzar en la exploración de la estructura de la materia no habremos alcanzado ya niveles en los que la «materialidad» –esto es, la materia– se desvanece transformándose en otra cosa diferente. Y ¿en qué otra cosa? Pues si estamos hablando de partículas que se manifiestan como vibraciones de cuerdas, ¿no será esa «otra cosa», una estructura matemática? Una vibración es, al fin y al cabo, la oscilación de algo material, pero en cuanto a estructura permanente tiene probablemente más de ente matemático que de ente material. Si fuese así, podríamos decir que se habría visto cumplido el sueño, o uno de los sueños, de Pitágoras. Los físicos habrían estado laborando duramente a lo largo de siglos, milenios incluso, para descubrir, finalmente, que la materia se les escapa de las manos, como si de una red se tratase, convirtiéndose en matemática, en estructuras matemáticas. La teoría de cuerdas, en resumen, resucita viejos problemas, acaso fantasmas. Problemas como el de la relación entre la física (y el mundo) y la matemática.

Independientemente de estos aspectos de naturaleza en esencia filosófica, hay otros que es imprescindible mencionar. Y es que hasta la fecha las teorías de cuerdas han demostrado muy poco, sobre todo si no olvidamos que la ciencia es explicación teórica, sí, pero también

experimentos, someter la teoría al juez último que es la comprobación experimental. Las teorías de cuerdas son admiradas por algunos, comentadas por muchos y criticadas por bastantes, que insisten en que su naturaleza es excesivamente especulativa. Así, Lee Smolin (2007, 17-18), un distinguido físico teórico, ha escrito en un libro dedicado a estas teorías:

En los últimos veinte años, se ha invertido mucho esfuerzo en la teoría de cuerdas, pero todavía desconocemos si es cierta o no. Incluso después de todo el trabajo realizado, la teoría no ha proporcionado ninguna predicción que pueda ser comprobada mediante experimentos actuales o, al menos, experimentos que podamos concebir en la actualidad. Las escasas predicciones limpias que propone ya han sido formuladas por otras teorías aceptadas.

Parte de la razón por la que la teoría de cuerdas no realiza nuevas predicciones es que parece presentarse en un número infinito de versiones. Aun limitándonos a las teorías que coinciden con algunos de los hechos básicos observados sobre nuestro Universo, por ejemplo, su vasto tamaño o la existencia de energía oscura, nos siguen quedando algo así como 10^{500} teorías de cuerdas diferentes; es decir, un 1 con 500 ceros detrás, más que todos los átomos conocidos del Universo. Una cantidad tal de teorías, nos deja poca esperanza de poder identificar algún resultado de algún experimento que no se pudiera incluir en alguna de ellas. Por tanto, no importa lo que muestren los experimentos, pues no se puede demostrar que la teoría de cuerdas sea falsa, aunque lo contrario también es cierto: ningún experimento puede demostrar que sea cierta.

Recordemos en este punto que uno de los sistemas metodológicos de la ciencia más influyentes continúa siendo el construido por el filósofo de origen austriaco, instalado finalmente en la London School of Economics de Londres, Karl Popper (1902-1994), quien siempre insistió en que una teoría que no es refutable mediante ningún experimento imaginable no es científica; esto es, que si no es posible imaginar algún experimento cuyos resultados contradigan las predicciones de una teoría, ésta no es realmente científica. Y aunque en mi opinión tal criterio es demasiado exigente para ser siempre verdad, constituye una buena guía. En cualquier caso, el futuro tendrá la última palabra sobre las teorías de cuerdas.

Nucleosíntesis estelar

En las páginas anteriores me he ocupado de los aspectos más básicos de la estructura de la materia, pero la ciencia no se reduce a buscar lo más fundamental, la estructura más pequeña; también pretende comprender aquello que nos es más próximo, más familiar. Es obligado, en este sentido, referirse a otro de los grandes logros de la física del siglo xx: la reconstrucción teórica de los procesos –nucleosíntesis– que condujeron a formar los átomos que encontramos en la naturaleza, y de los que nosotros mismos estamos formados. De estas cuestiones se ocupa la física nuclear, una disciplina relacionada, naturalmente, con la física de altas energías, aunque ésta sea más «fundamental» al ocuparse de estructuras más básicas que los núcleos atómicos.

24

No hay consenso acerca del porqué de la letra M. Unos opinan que es por Teoría Madre, otros Teoría del Misterio, Teoría de la Membrana o Teoría Matriz.

De hecho, la física de altas energías proporciona las bases sobre las que se asienta el edificio de la física nuclear que se ocupa de la nucleosíntesis estelar. Han sido, en efecto, los físicos de altas energías los que se han ocupado, y ocupan, de explicar cómo de la «sopa» indiferenciada de radiación y energía que surgió del Big Bang fueron formándose las partículas que constituyen los átomos.²⁵

Al ir disminuyendo la temperatura del Universo, esa sopa se fue diferenciando. A la temperatura de unos 30.000 millones de grados Kelvin (que se alcanzó en aproximadamente 0,11 segundos), los fotones –esto es, recordemos, la luz– se independizaron de la materia, distribuyéndose uniformemente por el espacio. Únicamente cuando la temperatura del Universo bajó a los 3.000 millones de grados Kelvin (casi 14 segundos después del estallido inicial) comenzaron a formarse –mediante la unión de protones y neutrones– algunos núcleos estables, básicamente el hidrógeno (un protón en torno al cual orbita un electrón) y el helio (dos protones y dos neutrones en el núcleo, con dos electrones como «satélites»). Estos dos elementos, los más ligeros que existen en la naturaleza, fueron –junto a fotones y neutrinos–, los principales productos del Big Bang, y representan aproximadamente el 73% (el hidrógeno) y el 25% (el helio) de la composición del universo.²⁶

Tenemos, por consiguiente, que el Big Bang surtió generosamente al Universo de hidrógeno y de helio. Pero ¿y los restantes elementos? Porque sabemos que hay muchos más elementos en la naturaleza. No hace falta ser un experto para saber que existe el oxígeno, el hierro, el nitrógeno, el carbono, el plomo, el sodio, el cinc, el oro y muchos elementos más. ¿Cómo se formaron?

Antes incluso de que los físicos de altas energías hubiesen comenzado a estudiar la nucleosíntesis primordial, hubo físicos nucleares que durante la primera mitad del siglo xx se ocuparon del problema de la formación de los elementos más allá del hidrógeno y el helio. Entre ellos es obligado mencionar a Carl Friedrich von Weizsäcker (1912-2007) en Alemania y a Hans Bethe (1906-2005) en Estados Unidos (Weizsäcker 1938; Bethe y Critchfield 1938; Bethe 1939a, b)²⁷. Justo cuando iba a comenzar la segunda mitad de la centuria, George Gamow (1904-1968) y sus colaboradores, Ralph Alpher (1921-2007) y Robert Herman (1914-1997), dieron otro paso importante, que fue seguido diecisiete años después por Robert Wagoner (n. 1938), William Fowler (1911-1995) y Fred Hoyle, que armados con un conjunto mucho más completo de datos de reacciones nucleares, explicaron que en el universo el litio constituye una pequeña fracción (10^{-8}) de la masa correspondiente al hidrógeno y al helio, mientras que el total de los restantes elementos representa un mero 10^{-11} (Wagoner, Fowler y Hoyle 1967).²⁸

Gracias a aportaciones como éstas –y las de muchos otros– ha sido posible reconstruir las reacciones nucleares más importantes en la nucleosíntesis estelar. Una de estas reacciones es la siguiente: dos núcleos de helio chocan y forman un átomo de berilio, elemento que ocupa el cuar-

to lugar (número atómico) en la tabla periódica, tras el hidrógeno, helio y litio (su peso atómico es 9, frente a 1 para el hidrógeno, 4 para helio y 6 para el litio). En realidad se produce más de un tipo de berilio; uno de ellos, el isótopo de peso atómico 8 es muy radiactivo, existiendo durante apenas una diezmilbillonésima de segundo, tras lo cual se desintegra produciendo de nuevo dos núcleos de helio. Pero si durante ese instante de vida el berilio choca con un tercer núcleo de helio puede formar un núcleo de carbono (número atómico 6 y peso atómico 12), que es estable. Y si las temperaturas son suficientemente elevadas, los núcleos de carbono se combinan y desintegran de maneras muy diversas, dando lugar a elementos como magnesio (número atómico 12), sodio (11), neón (10) y oxígeno (8). A su vez, los núcleos de oxígeno pueden unirse y formar azufre y fósforo. De este modo, se fabrican elementos cada vez más pesados. Hasta llegar al hierro (26).

Hechos de este tipo nos llevan a otra cuestión: la de cómo han llegado estos elementos a la Tierra, puesto que el lugar en el que fueron fabricados necesita de energías y temperaturas que no se dan en nuestro planeta. Y si suponemos que no deben existir diferencias grandísimas entre nuestro planeta y los demás –salvo en detalles como composición o posibilidad de que exista vida–, cómo han llegado a cualquier otro planeta. Pues bien, una parte de los elementos (hasta el hierro) que no se produjeron en los primeros instantes del Universo, se han fabricado sobre todo en el interior de estrellas. La emisión al espacio exterior de esos elementos puede tener lugar de tres maneras: mediante la lenta pérdida de masa en estrellas viejas, en la denominada fase «gigante» de la evolución estelar; durante los relativamente frecuentes estallidos estelares que los astrónomos denominan «novas»; y en las dramáticas y espectaculares explosiones que se producen en la etapa estelar final que alumbran las denominadas «supernovas» (una de estas explosiones fue detectada en 1987: la supernova SN1987A; la explosión había tenido lugar 170.000 años antes, el tiempo que ha tardado la luz en llegar a la Tierra).

Es sobre todo en la explosión de las supernovas cuando los elementos pesados fabricados en la nucleosíntesis estelar se difunden por el espacio. No se conoce demasiado bien por qué se producen estas explosiones, pero se cree que además de expulsar los elementos que acumulaba la estrella en su interior (salvo parte que retiene convertidos en objetos muy peculiares, como estrellas de neutrones), en el estallido se sintetizan elementos todavía más pesados que el hierro; elementos como el cobre, cinc, rubidio, plata, osmio, uranio, y así hasta una parte importante de los más de cien elementos que contiene en la actualidad la tabla periódica, y que son relativamente abundantes en sistemas estelares como el nuestro, el Sistema Solar.

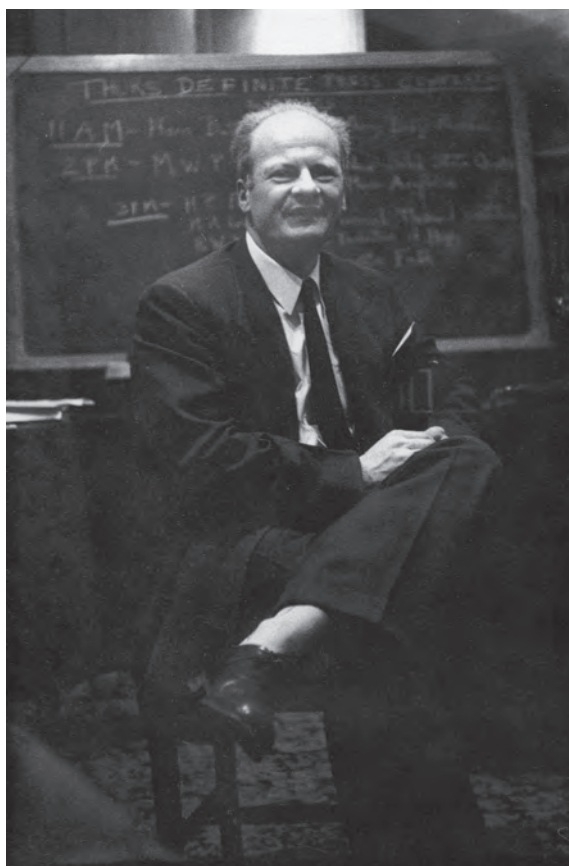
Precisamente por esta abundancia de elementos pesados, parece razonable pensar que el Sol es una estrella de segunda generación, formada, algo menos de hace 5.000 millones de años, por la condensación de residuos de una

25
Una magnífica y pionera exposición es la de Weinberg (1979).

26
No he tenido ocasión de mencionarlo: los neutrinos, que durante mucho tiempo se supuso no tenían masa (como los fotones), sí la tienen. Éste es otro de los resultados importantes de la física de la segunda mitad del siglo xx.

27
Por estos trabajos, Bethe recibió el Premio Nobel de Física de 1967.

28
Fowler obtuvo el Premio Nobel de Física de 1983 por estos trabajos, compartido con Chandrasekhar. Es sorprendente que Hoyle, que inició muchos de estos trabajos, quedase al margen de la decisión de la Academia sueca.



Hans Bethe (1957).

estrella anterior que murió en una explosión de supernova. El material procedente de semejante explosión se agrupó en un disco de gas y polvo con una protoestrella en el centro. El Sol se «encendió» cuando el núcleo central de materia se comprimió tanto que los átomos de hidrógeno se fundieron entre sí. Y alrededor suyo, a lo largo de bandas elípticas, siguiendo un proceso parecido pero menos intenso gravitacionalmente, se formaron los planetas de lo que llamamos Sistema Solar: Mercurio, Venus, la Tierra, Marte, Júpiter, Saturno, Urano, Neptuno y Plutón (aunque éste no es desde hace poco considerado en la categoría de «planeta»), y los satélites de éstos, como la Luna.

Desde esta perspectiva, la Tierra (formada hace unos 4.500 millones de años), al igual que los demás planetas, es algo parecido a un pequeño basurero (o cementerio) cósmico; un lugar en el que se han acumulado restos de la vida de estrellas, no lo suficientemente importantes como para dar lugar a un astro; esto es, aglomerados de elementos en cantidades tan pequeñas que no han podido desencadenar en su interior reacciones termonucleares como las que se producen en las estrellas. Pero al igual que en los basureros también se abre camino la vida, así ocurrió en esta Tierra nuestra, con su diámetro de, aproximadamente, 12.700 kilómetros y su peso de unas $6 \cdot 10^{21}$ (6 seguido de 21 ceros) toneladas. Nosotros somos testigos y demostración de este fenómeno.

Dentro de unos 7.500 millones de años, la zona central del Sol, en la que el hidrógeno se convierte en helio, aumentará de tamaño a medida que el hidrógeno se vaya consumiendo. Y cuando ese núcleo de helio alcance un tamaño suficiente, el Sol se dilatará hasta convertirse en una estrella de las denominadas gigantes rojas. Se hará tan gigantesca que su diámetro terminará alcanzando la órbita de la Tierra, acabando con ella. Antes de que suceda esto, la superficie terrestre llegará a estar tan caliente como para que el plomo se funda, hiervan los océanos y desaparezca todo rastro de vida. De esta manera, los procesos nucleares que nos dieron la vida acabarán con ella.

Más allá del mundo microscópico

Las teorías físicas de las que he estado tratando en las secciones precedentes son, es cierto, teorías cuánticas; ahora bien, el mundo de la física cuántica no se restringe a ellas y constituiría un grave error no referirse a otras novedades que surgieron en ese mundo durante la segunda mitad del siglo xx. Enfrentado con la difícil cuestión de buscar los avances más importantes, he optado por seleccionar dos grupos. El primero incluye desarrollos que han reforzado a la física cuántica frente a críticas como las que lideró Einstein junto a Podolsky y Rosen. El segundo trata de los trabajos que han puesto de relieve la existencia de fenómenos cuánticos *macroscópicos*.

Una teoría no local: el entrelazamiento cuántico

El fin de la ciencia es suministrar sistemas teóricos que permitan relacionar cuantos más fenómenos naturales mejor y que estos sistemas tengan capacidad predictiva. A esto le llamamos «explicar la naturaleza». Ahora bien, «explicar» no quiere decir encontrar respuestas que nos resulten familiares, que no violenten nuestras categorías explicativas más comunes: ¿por qué la naturaleza iba a seguir tales pautas? Ya aludí antes a que la física cuántica muestra con especial virulencia que la realidad nos puede resultar, de acuerdo con algunas teorías de gran éxito, profundamente contraintuitiva. Si este rasgo estaba claro desde que en 1925-1926 se contó con la mecánica cuántica, ahora lo está aún más. Veamos a qué me refiero.

En 1935, Albert Einstein, junto a dos colaboradores suyos, Boris Podolsky (1896-1966) y Nathan Rosen (1910-1995), publicaron un artículo (Einstein, Podolsky y Rosen 1935) en el que argumentaban que la mecánica cuántica no podía ser una teoría completa, que era necesario introducir nuevas variables. Sería largo explicar los argumentos que emplearon, que van más allá de lo puramente físico, adentrándose en terrenos claramente filosóficos (daban una definición de lo que es la «realidad física»); simplemente diré que su análisis condujo a que un físico natural de Belfast que trabajaba en el División de Teoría del CERN, John Stewart Bell (1928-1990), demostrase que existían unas relaciones (desigualdades) que se podían emplear para decidir experimentalmente qué tipo de teoría era

correcta, si una «completa» (que incluyese unas variables «ocultas» para la formulación cuántica) que obedeciese a los requisitos que Einstein, Podolsky y Rosen habían planteado en 1935 o la mecánica cuántica tradicional (Bell 1964, 1966). Provistos del análisis de Bell, John Clauser, Michael Horne, Abner Shimony y Richard Holt (1969) propusieron un experimento concreto para aplicar en él la prueba de las desigualdades de Bell. Este experimento se llevó a cabo en el Instituto de Óptica Teórica y Aplicada de Orsay, en las cercanías de París, por un equipo dirigido por Alain Aspect (n. 1947). Y el resultado (Aspect, Dalibard y Roger 1982) favoreció a la mecánica cuántica. Será rara, contraintuitiva, con variables que no se pueden determinar simultáneamente, socavará nuestra idea tradicional de lo que es la realidad, pero es cierta. El análisis de Bell y el experimento de equipo de Aspect mostraron además un rasgo de la mecánica cuántica que aunque conocido apenas había sido destacado: su no localidad; que todos los elementos de un sistema cuántico están conectados, *entrelazados*, entre sí; no importa que, alejados por tanta distancia, no sea posible que se transmita la señal de lo que le ha sucedido a uno de sus elementos a otro con la velocidad de la luz, la máxima permitida por la relatividad especial. En otras palabras, un elemento se «entera» –y reacciona– instantáneamente de lo que le sucede a otro independientemente de la distancia que les separe. La no localidad –que Einstein siempre rechazó, como contraria al «sentido común» físico– plantea, no hay duda, un problema de compatibilidad con la relatividad especial, pero no existe ninguna razón para pensar que no se encuentre en el futuro una generalización de la mecánica cuántica que resuelva esta dificultad. Eso sí, seguro que no será fácil.



John Bardeen.

La no localidad abre, asimismo, posibilidades que parecen pertenecer al dominio de la ciencia-ficción. Utilizaré, en este sentido, lo que ha escrito el divulgador científico Amir Aczel (2004, 20): «Mediante el entrelazamiento puede también "teleportarse" el estado de una partícula hasta un destino lejano, como sucede con el capitán Kirk en la serie televisiva *Star Trek* cuando pide ser proyectado de vuelta al *Enterprise*. Para ser preciso, nadie ha sido todavía capaz de teleportar a una persona. Pero el estado de un sistema cuántico ha sido teleportado en el laboratorio. Es más, este increíble fenómeno está comenzando a emplearse en criptografía y (podría usarse) en la futura computación cuántica».

Ideas –y, al menos en parte, realidades– como éstas, muestran que la ciencia puede superar incluso a la ciencia-ficción. En cualquier caso, consecuencias de la física cuántica como éstas pertenecerán más al siglo XXI que al que hace poco se cerró.

Fenómenos cuánticos macroscópicos:

«Lo submicroscópico deviene macroscópico»

Estamos acostumbrados a pensar que el dominio de la física cuántica es únicamente el microscópico, el de partículas elementales, átomos o radiaciones, pero no es así aunque es cierto que históricamente estos fenómenos fueron los responsables de la génesis de las teorías cuánticas. Las dos manifestaciones principales de esa física cuántica macroscópica son los condensados de Bose-Einstein y la superconductividad.

Condensados de Bose-Einstein

Desde el punto de vista de la teoría, los condensados (o condensación) de Bose-Einstein proceden del artículo que publicó en 1924 el físico hindú Satyendranath Bose (1894-1974), en el que introducía una nueva estadística (una forma de contar fotones) para explicar la ley de radiación de un cuerpo negro que había llevado a Max Planck a introducir la primera noción de cuantización en 1900. Fue Einstein quien reconoció el valor (y ayudó a que fuese publicado) del trabajo de Bose (1924), al que completó con dos artículos (Einstein 1924, 1925) en los que ampliaba las conclusiones de Bose. Señaló, por ejemplo, que se podría producir en un gas de fotones una condensación: «Una parte "se condensa" y la restante continúa siendo un gas perfecto saturado» (Einstein 1925). Con la expresión «condensación», Einstein quería decir que un grupo de fotones actúa como si fuese una unidad, sin que entre ellos parezca que existen fuerzas de interacción. Además, predijo que «si la temperatura desciende lo suficiente», se produciría en ese gas «una caída brutal y acelerada de la viscosidad en el entorno de una cierta temperatura», que estimaba para el helio líquido –en el que ya había indicios de tal superfluidez– en unos 2°K.

Hubo, no obstante, que esperar hasta el 8 de enero de 1938 para que se produjera un avance en la predicción einsteiniana de la existencia de superfluidez. Fue entonces

cuando se publicaron en la revista inglesa *Nature* dos breves artículos, uno a cargo del Piotr Kapitza (1894-1984), director del Instituto de Problemas Físicos en Moscú y anteriormente (hasta que en 1934 Stalin le retuvo en la Unión Soviética, durante uno de sus viajes de vacaciones) catedrático en el Laboratorio Cavendish de Cambridge, y otro de dos jóvenes físicos canadienses que estaban trabajando en el Laboratorio Mond que la Royal Society patrocinaba en Cambridge, Jack Allen (1908-2001) y Don Misener (1911-1996). En ellos (Kapitza 1938; Allen y Misener 1938) se anunciaba que el helio líquido fluía casi sin sufrir la resistencia de la viscosidad por debajo de 2,18°K. Fueron, sin embargo, Fritz London (1900-1954) y Laszlo Tisza (n. 1907) quienes demostraron teóricamente que este fenómeno constituía la evidencia de la superfluidez.²⁹

Se trataba, por supuesto, de la vieja idea que Einstein había propuesto en 1924, y a la que apenas se había prestado atención, aunque ahora más desarrollada y aplicada a otros sistemas muy diferentes a los gases ideales considerados por el creador de la relatividad.

Es preciso señalar, sin embargo, que a pesar de que en la actualidad damos a los descubrimientos de 1938 un gran valor, en la medida en que mostraban macroscópicamente un comportamiento cuántico, en su momento este aspecto no se destacó tanto. Para comprender mejor la relación entre la condensación de Bose-Einstein y los aspectos macroscópicos de la física cuántica, hubo que tratar con átomos, producir «superátomos», conjuntos de átomos que se comportasen como una unidad y fuesen perceptibles macroscópicamente. Semejante logro se alcanzó mucho más tarde, en 1995. Aquel año, dos físicos de Colorado, Eric Cornell (n. 1961) y Carl Wieman (n. 1951), produjeron un superátomo de rubidio, y unos meses después Wolfgang Ketterle (n. 1957), del MIT, otro de sodio (los tres recibieron el Premio Nobel de Física de 2001). Veamos cómo han descrito los dos primeros su aportación (Cornell y Wieman 2003, 82):

Nuestro grupo de investigación del Instituto Conjunto de Astrofísica de Laboratorio (o JILA ahora), en Boulder, creó en junio de 1995 una gota, aunque minúscula, maravillosa. Al enfriar 2.000 átomos de rubidio hasta una temperatura de menos de 100 milmillonésimas de grado sobre el cero absoluto (100 milmillonésimas de grados kelvin) hicimos que los átomos perdiesen su identidad individual y se comportaran como si fuesen un solo «superátomo». Las propiedades físicas de todos ellos, sus movimientos, por ejemplo, se volvieron idénticas. Este condensado de Bose-Einstein, el primero observado en un gas, viene a ser el análogo material del láser, con la diferencia de que en el condensado son átomos, no fotones, los que danzan al unísono.³⁰

Y más adelante añadían (Cornell y Wieman 2003, 82-84):

Raras veces vemos los efectos de la mecánica cuántica reflejados en la conducta de una cantidad macroscópica de materia. Las contribuciones incoherentes del inmenso número de partículas de cualquier porción de materia oscurecen la naturaleza ondulatoria de la mecánica cuántica; sólo podemos inferir sus efectos. Pero en la condensación de Bose la naturaleza ondulatoria de cada átomo está en fase con las demás; y lo está

de manera precisa. Las ondas mecanocuánticas atraviesan la muestra entera y se observan a simple vista. Lo submicroscópico deviene macroscópico.

La creación de condensados de Bose-Einstein ha arrojado luz sobre viejas paradojas de la mecánica cuántica. Por ejemplo, si dos o más átomos están en un solo estado mecanocuántico, y eso es lo que pasa en un condensado, será imposible distinguirlos, se haga la medición que se haga. Los dos átomos ocuparán el mismo volumen de espacio, se moverán a la misma velocidad, dispersarán luz del mismo color, etc.

En nuestra experiencia, basada en el trato constante de la materia a temperaturas normales, no hay nada que nos ayude a comprender esta paradoja. Por un motivo: a las temperaturas normales y a las escalas de magnitud en que nos desenvolvemos, es posible describir la posición y el movimiento de todos y cada uno de los objetos de un conjunto... A temperaturas bajísimas o a escalas de magnitud pequeñas, la mecánica clásica va perdiendo vigor... No podemos saber la posición exacta de cada átomo, y es mejor imaginarlos como manchas imprecisas. La mancha es un paquete de ondas, la región del espacio donde cabe esperar que esté el átomo. Conforme el conjunto de átomos se enfría, crece el tamaño de los paquetes de ondas. Mientras cada uno esté espacialmente separado de los demás será posible, al menos en principio, distinguir átomos entre sí. Pero cuando la temperatura llega a ser lo bastante baja los paquetes de ondas de los átomos vecinos se solapan. Entonces, los átomos «se Bose-condensarán» en el menor estado de energía que sea posible, y los paquetes de ondas se fundirán en un solo paquete macroscópico. Los átomos sufrirán una crisis cuántica de identidad: ya no podremos distinguir unos de los otros.

Superconductividad

La superconductividad es otro de los fenómenos físicos en los que la cuantización se manifiesta macroscópicamente. El fenómeno en sí fue descubierto hace mucho, en 1911, cuando el físico holandés Heike Kamerlingh Onnes (1853-1926), el gran pionero y experto mundial en bajas temperaturas, encontró en su laboratorio de Leiden que cuando enfriaba el mercurio metal a 4°K, se anulaba por completo su resistencia ante el paso de una corriente eléctrica (Kamerlingh Onnes 1911). Una vez iniciada esa corriente, continuaría indefinidamente aunque no se le aplicase ninguna diferencia de potencial. Más tarde se encontró que otros metales y compuestos se hacían también superconductores a temperaturas cercanas al cero absoluto de temperatura. Ahora bien, una cosa es la evidencia experimental y otra la explicación teórica, y ésta tardó en llegar. Fue, en efecto, en 1957 cuando los estadounidenses John Bardeen (1908-1991), Leon Cooper (n. 1930) y John Robert Schrieffer (n. 1931) dieron con tal teoría (conocida por las iniciales de sus apellidos: BCS).³¹ Su explicación (Bardeen, Cooper y Schrieffer 1957) consistía en que a partir de una cierta temperatura los electrones que transportan la corriente eléctrica en un elemento o compuesto superconductor se agrupan en parejas –como había supuesto con anterioridad Cooper (1956); de ahí los «pares de Cooper»– que actúan como bosones; esto es, partículas como los fotones que no están sometidos a ciertos requisitos cuánticos. Este agrupamiento se produce a temperaturas muy bajas y se debe a la interacción entre los electrones y la red de átomos metálicos del compuesto superconductor. En este momento, agrupados, los pares de

29 London (1938), Tisza (1938).

30 El cero absoluto de temperatura (0°K) corresponde a -273,15°C. A esa temperatura, las moléculas no se mueven.

31 Los tres compartieron el Premio Nobel de Física en 1972.

electrones marchan como un armonioso ejército de bosones que ignoran los impedimentos atómicos. Es así como se manifiesta macroscópicamente este efecto cuántico.

La teoría BCS constituyó un éxito formidable de la física cuántica, pero no es totalmente satisfactoria, como se puso en evidencia por su incapacidad de predecir la existencia de superconductividad en materiales cerámicos a temperaturas mucho más elevadas que las que se manejaban hasta entonces. Fue en 1986, en los laboratorios de IBM de Zúrich, donde Georg Bednorz (n. 1950) y Alexander Müller (n. 1927) encontraron que un óxido de lantano, bario y cobre era superconductor a temperaturas tan altas (no, claro, para nuestras experiencias cotidianas) como 35°K.³² El año siguiente, Paul Chu (1987) elevó la escala de temperaturas superconductoras, hallando un óxido de itrio, bario y cobre que se volvía superconductor a la temperatura de 93°K, una temperatura que se puede alcanzar sin más que bañar ese óxido en nitrógeno líquido que –a diferencia del helio– es abundante y barato. Desde entonces, el número de estos materiales y de las temperaturas a las que se hacen superconductores no ha hecho sino aumentar.

El hallazgo de Bednorz y Müller (1986),³³ por el que recibieron el Premio Nobel de Física en 1987, abre nuevas perspectivas no sólo a la física sino también, y acaso sobre todo, a la tecnología: materiales superconductores a temperaturas a las que se puede llegar en el mundo cotidiano (esto es, fuera del laboratorio) pueden tal vez revolucionar nuestras vidas algún día.

Artifugios cuánticos: transistores, chips, máseres y láseres

El comentario anterior, la relevancia de la física cuántica en la tecnología, va mucho más allá de la superconductividad. Es posible que materiales superconductores cambien en el futuro nuestras vidas, pero de lo que no hay duda es

32

Hasta entonces, el superconductor de temperatura más elevada era una aleación de niobio y germanio, que actuaba así a 23°K.

33

Ver, asimismo, Müller y Bednorz (1987).

34

Como su nombre indica, y aunque ciertamente no es una definición que en principio ilustre mucho, un semiconductor es un material que conduce la electricidad en un rango entre el que caracteriza a un metal y a un aislante. Habitualmente, la conductividad de los semiconductores puede ser mejorada añadiéndoles pequeñas impurezas o por otros factores, como en el caso del silicio, que es muy poco conductor a bajas temperaturas, aunque su conductividad aumenta si se aplica calor, luz o una diferencia de potencial (por ello el silicio se utiliza en transistores, rectificadores y circuitos integrados).

35

Ver Shockley (1947, 1948) y Bardeen y Brattain (1948, 1949).

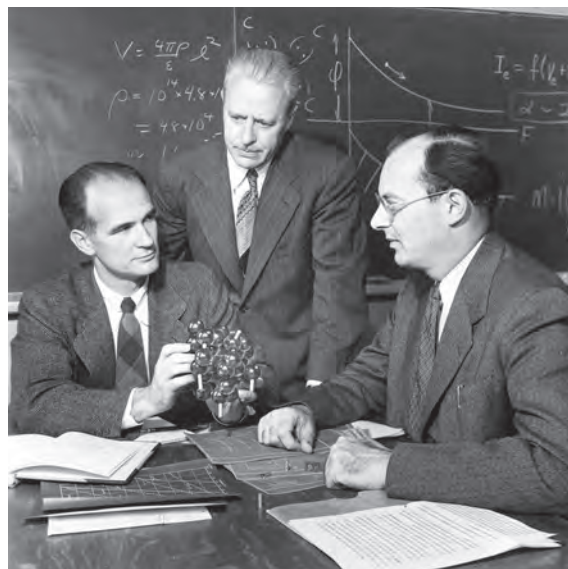
de otros materiales, los semiconductores, ya las han cambiado.³⁴ La primera gran aplicación de los semiconductores llegó con la invención del transistor, debida a John Bardeen, William Shockley (1910-1989) y Walter Brattain (1902-1987), mientras trabajaban en el departamento de física del estado sólido de los laboratorios Bell.³⁵ En 1956 los tres recibieron el Premio Nobel de Física, el primero de los dos que ganó Bardeen (el segundo, como ya vimos, fue por la superconductividad).

Un transistor es un dispositivo electrónico hecho de material semiconductor, que puede regular una corriente que pasa a través de él y también actuar como amplificador o célula fotoeléctrica, y, que comparado con los tubos de vacío que le precedieron, necesita cantidades muy pequeñas de energía para funcionar; además son más estables, compactos, actúan instantáneamente y duran más.

Tras los transistores vinieron los circuitos integrados, minúsculos y muy delgados dispositivos en los que se fundamenta el mundo digital. Los circuitos integrados se fabrican sobre un sustrato (habitualmente de silicio) depositando finas películas de materiales que, ora conducen, ora aíslan, la electricidad. Estas películas, ensambladas según patrones elaborados de antemano, forman transistores (cada circuito integrado puede albergar millones de transistores) que funcionan como interruptores encargados de controlar el flujo de electricidad a través del circuito, o *chip*.

Integrados en los *chips*, los transistores desempeñan funciones básicas en los billones y billones de microprocesadores que, repartidos, controlan, por ejemplo, motores de coche, teléfonos celulares, misiles, satélites, redes de gas, hornos microondas, ordenadores o aparatos para discos compactos. Han cambiado, literalmente, las formas en las que nos comunicamos, relacionamos con el dinero, escuchamos música, vemos televisión, conducimos coches, lavamos nuestras ropas o cocinamos.

Hasta la llegada de los transistores y circuitos integrados, las máquinas de calcular utilizadas eran gigantescos amasijos de componentes electrónicos. Durante la Segunda Guerra Mundial se construyó una de las primeras máquinas de calcular electrónicas: el Electronic Numerical Integrator and Computer (Computador Integrador Numérico Electrónico, también conocido por sus siglas inglesas, ENIAC). Tenía 17.000 tubos electrónicos, unidos por miles de cables, pesaba 30 toneladas y consumía 174 kilowatios. Podemos considerarlo el paradigma de la primera generación de computadores. Con los transistores llegó, en la década de 1950, la segunda generación: el primer computador surgido de la física del estado sólido –una rama de la física cuántica– fue el TRADIC (de *Transistor Digital Computer*), construido en 1954 por los laboratorios Bell para la Fuerza Aérea estadounidense; utilizaba 700 transistores y podía competir en velocidad con ENIAC. A finales de la década de 1960, gracias a los circuitos integrados, llegaría la tercera generación, a la que siguió una cuarta, con



Los inventores del transistor: William Shockley, Walter Brattain y John Bardeen.



Aleksandr Prokhorov y Nikolai Basov enseñando a Charles Townes (entre ambos) su laboratorio en Moscú (1965).

computadoras que utilizan microprocesadores y refinados lenguajes de programación. Y ya se habla de los computadores cuánticos, que en lugar de utilizar *bits*, que toman valores 1 o 0, definidos, recurren a *qubits*, *bits cuánticos*, que pueden estar en superposiciones cuánticas de 0 y 1, lo mismo que un fotón puede estar en superposiciones de polarización horizontal y vertical. Pero los computadores cuánticos, si se logran, pertenecerán a, posiblemente, la segunda mitad del siglo XXI.

Gracias a todos estos desarrollos nos encontramos sumergidos de lleno en un mundo pleno de computadoras que realizan, a velocidades y fiabilidades extraordinarias, todo tipo de funciones, y sin las cuales nuestra vida sería muy diferente. Y nada de esto, es muy importante destacarlo, se habría producido sin los resultados obtenidos en una rama de la física cuántica: la física del estado sólido (también denominada de la materia condensada).

En el haber de esta rama de la física se encuentra también el que ha estrechado las relaciones entre ciencia y sociedad. En 1955, por poner un ejemplo, Shockley, uno de los inventores del transistor, abandonó los laboratorios Bell para fundar su propia compañía en el área de la bahía de San Francisco. El Shockley Semiconductor Laboratory abrió sus puertas en febrero de 1956, reclutando un excelente grupo de profesionales. No tuvo, sin embargo, demasiado éxito, pero a la postre constituyó el germen que condujo al crecimiento de una zona en la que se agruparon diversas compañías tecnológicas en un lugar de California que terminó siendo conocido como Silicon Valley, el Valle del Silicio.

Ciencia y técnica se alían en este mundo tecnocientífico de una manera tan, digamos, íntima, que no es cierto que las innovaciones fundamentales se den sólo en los enclaves científicos y los negocios en los tecnológicos. Recordemos, en este sentido, que las técnicas fundamentales (proceso «planar») para la fabricación de los *chips* fueron ideadas en 1957 por Jean Hoerni (1924-1997), de

la empresa Fairchild Semiconductors. El primer circuito integrado fue construido allí por Robert N. Noyce (1927-1990) en 1958. Diez años después (1968), Noyce dejó Fairchild para fundar, junto a Gordon Moore (n. 1929), Intel, donde dirigió con Ted Hoff (n. 1937), la invención del microprocesador, que generó una nueva revolución.

Quiero mencionar, asimismo, que el desarrollo de los microprocesadores electrónicos ha estimulado –y, a su vez, se ha visto él mismo beneficiado por– la denominada «nanotecnología», cuyo objetivo es el control y manipulación de la materia a una escala del orden de entre 1 y 100 nanómetros (1 nanómetro equivale a 10^{-9} metros). La nanotecnología es más una técnica (un conjunto de técnicas) que una ciencia, pero de ella cabe esperar (en parte ya lo está dando) desarrollos que no sólo beneficien nuestras posibilidades materiales sino también al conocimiento científico más básico.

Másers y láseres

Aún no he mencionado –aunque cronológicamente precedieron a algunos de los desarrollos anteriores– al máser y al láser, acrónimos de, respectivamente, *Microwave Amplification by Stimulated Emission of Radiation* (amplificación de microondas mediante emisión estimulada de radiación) y de *Light Amplification by Stimulated Emission of Radiation* (amplificación de luz por emisión estimulada de radiación).

Desde el punto de vista de la teoría, estos instrumentos–procedimientos para amplificar ondas de la misma frecuencia (longitud de onda) se explican en base al contenido de dos artículos de Einstein (1916a, b). Sin embargo, su realización práctica, con todos los nuevos elementos teóricos y experimentales que ello conllevó, no llegó hasta la década de 1950. Los responsables de este logro fueron, de manera independiente, los físicos del Instituto Lebedev de Física de Moscú, Aleksandr M. Prokhorov (1916-2002) y Nikolai G. Basov (1922-2001), y el estadounidense Charles Townes (n. 1915), de la Universidad de Columbia, Nueva York (los tres compartieron el Premio Nobel de Física de 1964).

En mayo de 1952, durante una conferencia sobre radio-espectroscopía en la Academia de Ciencias de la URSS, Basov y Prokhorov describieron el principio del máser, aunque no publicaron nada hasta dos años después (Basov y Prokhorov 1954). Y no sólo describieron su principio, sino que también Basov construyó uno como parte de su tesis doctoral, unos pocos meses después de que Townes hiciese lo propio.

Merece la pena resumir cómo Townes llegó por su parte a la misma idea del máser, ya que ilustra acerca de lo muy diversos que pueden ser los elementos que forman parte de los procesos de descubrimiento científico. Tras permanecer en los laboratorios Bell entre 1939 y 1947, en donde se ocupó, entre otros temas, de la investigación relacionada con el radar, Townes pasó al Radiation Laboratory de la Universidad de Columbia, creado durante la Segunda Guerra Mundial para desarrollar radares, esenciales para

el desarrollo de la guerra. Al igual que otras instituciones, este laboratorio continuó recibiendo dinero de los militares después de la contienda, dedicando el 80% de su presupuesto al desarrollo de tubos que generasen microondas. En la primavera de 1950, Townes organizó en Columbia para la Oficina de Investigación de la Marina un comité asesor para considerar nuevas formas de generar microondas de menos de un centímetro. Tras un año de considerar la cuestión, se le ocurrió un nuevo enfoque antes de asistir a una de las sesiones de su comité: era la idea de máser. Cuando logró, en 1954 y en colaboración con un joven doctor, Herbert J. Zeiger, y un doctorando, James P. Gordon, hacer realidad operacional esa idea utilizando un gas de moléculas de amoniaco (Gordon, Zeiger y Townes 1954), resultó que las oscilaciones producidas por el máser se caracterizaban no sólo por su alta frecuencia y potencia, sino también por su uniformidad. El máser, en efecto, produce una emisión coherente de microondas; esto es, radiación altamente concentrada, de una única longitud de onda.

Incluso antes de que los másers empezasen a proliferar, algunos físicos comenzaron a intentar extender su idea a otras longitudes de onda. Entre ellos se encontraba el propio Townes (también Basov y Prokhorov), quien a partir del otoño de 1957 inició trabajos para ir desde las microondas a la luz visible, colaborando con su cuñado, Arthur Schawlow (1921-1999), un físico de los laboratorios Bell. Fruto de sus esfuerzos fue un artículo básico, en el que mostraban cómo se podría construir un láser, al que todavía denominaban «máser óptico» (Schawlow y Townes 1958). No está de más mencionar que los abogados de los laboratorios Bell, para los que trabajaba Schawlow y con los que Townes tenía un contrato de asesor, pensaron que

la idea del láser no tenía interés suficiente como para ser patentada; únicamente lo hicieron ante la insistencia de Townes (Schawlow y Townes 1960).

La carrera por construir un láser se intensificó a partir de entonces. Aunque la historia posterior no siempre ha sido lo suficientemente clara en este punto, el primero que tuvo éxito fue Theodore Maiman (1927-2007), de los Hughes Research Laboratories de Malibu (California), que consiguió poner en funcionamiento un láser de rubí el 16 de mayo de 1960. Maiman envió a la entonces recién establecida *Physical Review Letters* un manuscrito con sus resultados, pero su editor, Samuel Goudsmit, lo rechazó como «sólo otro artículo sobre el máser». En su lugar, Maiman recurrió a *Nature*, en cuyo número del 6 de agosto de 1960 consiguió publicar el resultado de su trabajo (Maiman 1960). Poco después, Schawlow anunciaba –en *Physical Review Letters*– que había puesto en funcionamiento otro láser, también de rubí (considerablemente más grande y potente que el de Maiman) junto a cinco colaboradores (Collins, Nelson, Schawlow, Bond, Garret y Kaiser 1960). Habida cuenta de estos hechos es cuestionable que fuese Schawlow quien recibiese en 1981 el Premio Nobel (compartido con Nicolaas Bloembergen y con Kai Siegbahn) aunque formalmente fuese por su contribución (y la de Bloembergen) al desarrollo de la espectroscopia láser.³⁶

Los másers y, sobre todo, los láseres (otro «hijo», por cierto, de la física cuántica que muestra macroscópicamente efectos cuánticos) son instrumentos bien conocidos por el público, en particular algunas de sus aplicaciones (por ejemplo, en operaciones de desprendimientos de retina, en las que se emplean láseres); sin embargo, lo son menos otras de gran importancia científica. Una de ellas es su utilización en espectroscopia. Al ser radiaciones monocromáticas de gran energía es posible dirigir las con precisión a niveles atómicos determinados; los resultados obtenidos permiten conocer mucho mejor las propiedades de las moléculas, cuya estructura las hace mucho más complicadas de estudiar que los átomos.

Un mundo no lineal

Los descubrimientos y desarrollos a los que me he referido hasta ahora son, probablemente, los más sobresalientes desde un punto de vista, digamos, fundamental; no obstante, no incluyen un conjunto de avances que están abriendo nuevas y sorprendentes ventanas a la comprensión científica de la naturaleza. Se trata de los fenómenos no lineales; esto es, los gobernados por leyes que involucran ecuaciones con términos cuadráticos.³⁷

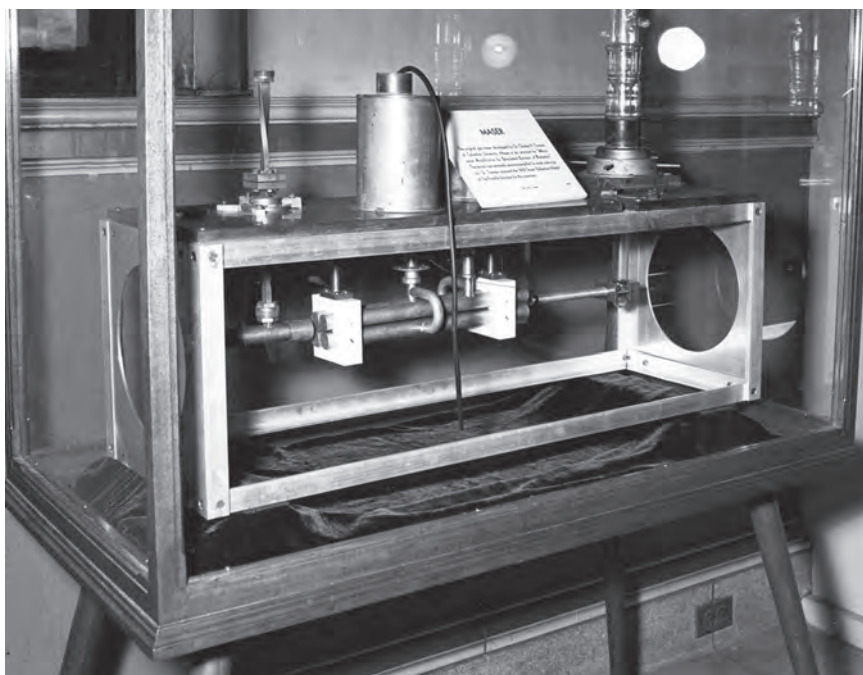
Si repasamos la historia de la física hasta bien entrado el siglo xx, nos encontramos con que sus teorías más básicas o son esencialmente lineales (los casos de la teoría de la gravitación universal de Newton y de la electrodinámica de Maxwell), o, permitiendo ser utilizadas para sistemas

36

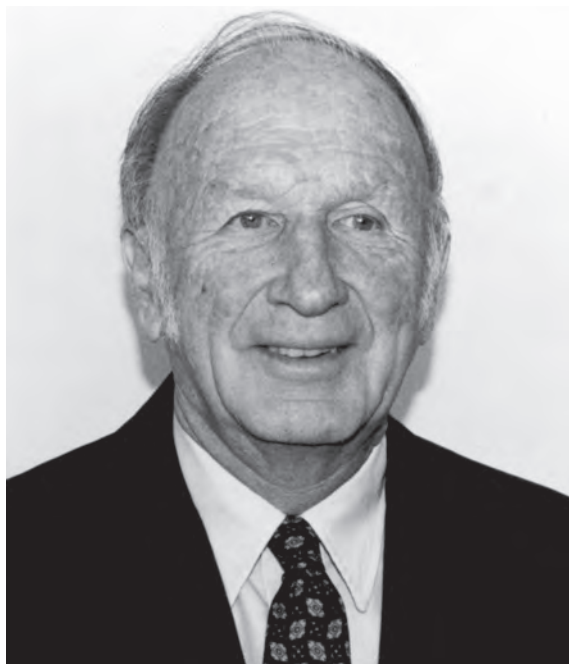
Siegbahn lo recibió por sus aportaciones al desarrollo de la espectroscopia electrónica de alta resolución.

37

De manera simbólica, se puede decir que la expresión de la linealidad es la ecuación $A + A = 2A$, mientras que en el mundo de la no linealidad, el universo en el que la reunión de dos seres genera, crea, nuevas propiedades, $A+A \neq 2A$. De manera rigurosa, esto es, matemáticamente, la diferencia esencial entre un sistema lineal y otro no lineal es que mientras que dos soluciones de un sistema lineal se pueden sumar formando una nueva solución del sistema inicial («principio de superposición»), esto no es cierto en el caso de un sistema no lineal.



El primer máser construido por Townes y sus colaboradores, expuesto en el Franklin Institute (Filadelfia).



Edward Lorenz.

no lineales, como sucede con la mecánica newtoniana, se han aplicado mayoritariamente a sistemas lineales, incluso cuando es transparentemente claro que ello implica una aproximación a la realidad. El ejemplo más sencillo en este sentido es el de un péndulo plano simple. Cualquier estudiante de bachillerato, no digamos de física, sabe que la ecuación diferencial que se utiliza para describir el movimiento de este tipo de péndulo es

$$d^2\theta(t)/dt^2 + (g/l)\theta(t) = 0$$

donde θ representa el desplazamiento angular del péndulo, l la longitud del péndulo, g la aceleración de la gravedad y t el tiempo. Ahora bien, cuando se deduce (no es un problema difícil) la ecuación que debe cumplir el movimiento de un péndulo plano simple, no es ésta la ecuación a la que se llega, sino

$$d^2\theta(t)/dt^2 + (g/l)\text{sen}\theta(t) = 0$$

que es obviamente no lineal, ya que

$$\text{sen}(\theta_1 + \theta_2) \neq \text{sen}\theta_1 + \text{sen}\theta_2$$

Para evitar esta circunstancia, que complica enormemente la resolución del problema, se restringe el problema a oscilaciones pequeñas, esto es, a ángulos pequeños, lo que permite utilizar el desarrollo en serie de Taylor de la función seno

$$\text{sen}\theta \approx \theta - \theta^3/6 + \dots$$

y quedándose únicamente con el primer término, obtener la primera (lineal) de las dos ecuaciones citadas.

Lo que este ejemplo tan sencillo nos muestra es que la denominada «física clásica» no es ajena a sistemas no lineales,

pero que trata de evitarlos debido a la dificultad matemática de su tratamiento: no existen, de hecho, métodos matemáticos generales sistemáticos para tratar las ecuaciones no lineales. Por supuesto, son numerosos los problemas conocidos de antiguo asociados a sistemas (leyes) no lineales, especialmente los pertenecientes al ámbito de la hidrodinámica, de la física de los fluidos. Así, por ejemplo, cuando el agua fluye con una velocidad pequeña por una tubería, su movimiento (denominado *laminar*), regular y predecible, se describe mediante ecuaciones lineales, pero cuando las velocidades involucradas son elevadas entonces el movimiento del agua se hace *turbulento*, formándose remolinos que siguen trayectorias irregulares, aparentemente erráticas, características típicas de un comportamiento no lineal. La aerodinámica, naturalmente, es otro ejemplo de dominio no lineal, como saben muy bien todos aquellos implicados en el diseño de aviones.³⁸

La riqueza de los sistemas no lineales es extraordinaria; la riqueza y las novedades que aportan con respecto a los lineales. Desde el punto de vista matemático (que con frecuencia encuentra su correlato en dominios reales), las ecuaciones-sistemas no lineales pueden mostrar transiciones de comportamientos regulares a aparentemente arbitrarios; pulsos localizados, que en sistemas lineales producen perturbaciones que decaen más pronto que tarde, mantienen su individualidad en los sistemas no lineales; esto es, dan lugar a estructuras localizadas y altamente coherentes, con las obvias implicaciones que este fenómeno puede tener en la aparición y mantenimiento de estructuras relacionadas con la vida (desde las células y organismos pluricelulares hasta incluso, aunque pueda parecer una idea peregrina, pensamientos). Uno de los primeros ejemplos conocidos de este tipo de comportamiento son los célebres «solitones», soluciones de la ecuación no lineal en derivadas parciales denominada de Korteweg-de Vries (o ecuación KdV), desarrollada en 1895 como una descripción aproximada de las ondas de agua que se movían en un canal estrecho y poco profundo. No fue, sin embargo, hasta 1965 cuando Norman Zabusky y Martin Kruskal encontraron una solución de esta ecuación que representa una de las formas más puras de estructura coherente en movimiento (Zabusky y Kruskal 1965): el solitón, una onda solitaria que se mueve con velocidad constante. Lejos de ser entequiras matemáticas, los solitones se manifiestan en la naturaleza: por ejemplo, en las ondas superficiales (que se mueven esencialmente en una dirección) observadas en el mar de Andamán, que separa las islas de Andamán y de Nicobar de la península de Malasia.

El caos

Un caso particularmente importante de sistema no lineal son los sistemas caóticos. Un sistema caótico se caracteriza porque las soluciones de las ecuaciones que lo representan son extremadamente sensibles a las condiciones iniciales; esto es, son tales que si se cambian

38

De las grandes teorías de la física clásica, la más intrínsecamente no lineal es la teoría de la relatividad general (las ecuaciones del campo de esta teoría de la interacción gravitacional son no lineales).

un poco, minúsculamente, esas condiciones, entonces la solución (la trayectoria que sigue el objeto descrito por la solución) se ve modificada radicalmente, siguiendo un camino completamente diferente, al contrario que en los sistemas no caóticos, aquellos con los que la física nos ha familiarizado durante siglos, en los que pequeños cambios en las condiciones iniciales no alteran sustancialmente la solución. Es por su extrema variabilidad frente a aparentemente insignificantes cambios en sus puntos y condiciones de partida, que esas soluciones y los sistemas a que pertenecen se denominan *caóticos*. Pero que sean «caóticos» no significa que no puedan ser sometidos a leyes expresables en términos matemáticos. Es preciso recalcar que los sistemas caóticos están descritos por leyes codificadas en expresiones matemáticas, expresiones de hecho similares a las que pueblan el universo de las leyes lineales de la dinámica newtoniana.

El tiempo meteorológico constituye uno de los grandes ejemplos de sistemas caóticos; de hecho, fue estudiándolo cuando se descubrió realmente el caos: pequeñas perturbaciones en la atmósfera pueden cambiar el clima en proporciones enormes. Su descubridor fue el meteorólogo teórico estadounidense Edward Norton Lorenz (1938-2008).

Estaba trabajando en sus investigaciones sobre el tiempo atmosférico, desarrollando modelos matemáticos simples cuyas propiedades exploraba con la ayuda de ordenadores, cuando, en 1960, observó que algo raro ocurría cuando repetía cálculos anteriores. He aquí como él mismo reconstruyó los acontecimientos y su reacción en un libro que escribió años después, *La esencia del caos* (Lorenz 1995, 137-139):

En un momento dado, decidí repetir algunos de los cálculos con el fin de examinar con mayor detalle lo que estaba ocurriendo. Detuve el ordenador, teclé una línea de números que había salido por la impresora un rato antes y lo puse en marcha otra vez. Me fui al vestíbulo a tomarme una taza de café y regresé al cabo de una hora, tiempo durante el cual el ordenador había simulado unos dos meses de tiempo meteorológico. Los números que salían por la impresora no tenían nada que ver con los anteriores. Inmediatamente pensé que se había estropeado alguna válvula o que el ordenador tenía alguna otra avería, cosa nada infrecuente, pero antes de llamar a los técnicos decidí comprobar dónde se encontraba la dificultad, sabiendo que de esa forma podría acelerar la reparación. En lugar de una interrupción brusca, me encontré con que los nuevos valores repetían los anteriores en un principio, pero que enseguida empezaban a diferir, en una, en varias unidades, en la última cifra decimal, luego en la anterior y luego en la anterior. La verdad es que las diferencias se duplicaban en tamaño más o menos constantemente cada cuatro días, hasta que cualquier parecido con las cifras originales desaparecía en algún momento del segundo mes. Con eso me bastó para comprender lo que ocurría: los números que yo había teclado no eran los números originales exactos sino los valores redondeados que había dado a la impresora en un principio. Los errores redondeados iniciales eran los culpables: se iban amplificando constantemente hasta dominar la solución. Dicho con terminología de hoy: se trataba del caos.

Lo que Lorenz había observado empíricamente, ayudado por su ordenador, es que existen sistemas que pueden des-

plegar un comportamiento impredecible (lo que no quiere decir «no sujeto a leyes»): pequeñas diferencias en una sola variable tienen efectos profundos en la historia posterior del sistema. Por eso, por ser un sistema caótico, el tiempo meteorológico es tan difícil de predecir, tan, como solemos decir, *imprevisible*.

El artículo en el que presentó sus resultados (Lorenz 1963) constituye uno de los grandes logros de las ciencias físicas del siglo xx, aunque pocos científicos que no fueran meteorólogos repararon entonces en él, una situación que cambiaría de forma radical a lo largo de las décadas siguientes. No poco tuvo que ver en semejante cambio de actitud la célebre frase «El aleteo de una mariposa en Brasil puede producir un tornado en Texas», que Lorenz incluyó en una conferencia que pronunció el 29 de diciembre de 1972 en una sesión de la reunión anual de la American Association for the Advancement of Science.³⁹

Cada vez está más claro que los fenómenos caóticos abundan en la naturaleza. Los encontramos ya en dominios propios de la economía, aerodinámica, la biología de poblaciones (en, por ejemplo, algunos modelos «presa-depredador»), termodinámica, química y, por supuesto, en el mundo de las ciencias biomédicas (un ejemplo es el de algunas arritmias). Parece que puede manifestarse incluso en los aparentemente estables movimientos planetarios.

Las consecuencias que para nuestra visión del mundo tiene el descubrimiento del caos y su, por lo que parece, ubicua presencia son incalculables. El mundo no es como suponíamos. Y no lo es sólo en los dominios atómicos, descritos por la física cuántica, en los que reinan la probabilidad y la incertidumbre, sino también en aquellos gobernados por las más «clásicas» leyes de tipo newtoniano. Newtonianas, sí, pero no como las que utilizó el gran Isaac Newton y todos sus seguidores, esto es, leyes *lineales* sino *no lineales*. La naturaleza no es lineal, sino no lineal. No todos los sistemas no lineales son caóticos, pero sí que todos los sistemas caóticos son no lineales. El mundo es por ello más complicado de explicar; no podemos predecir todo lo que va a suceder siguiendo el viejo, newtoniano, estilo, pero ¿por qué iba a ser la naturaleza tan «sencilla»? Lo maravilloso es que seamos capaces de descubrir tales comportamientos y las leyes matemáticas que subyacen en ellos.

Podría, y acaso debería, haber mencionado otros desarrollos que se han producido, o iniciado, durante la segunda mitad del siglo xx, como, por ejemplo, en la termodinámica del no equilibrio, en la que uno de los elementos centrales son los gradientes, diferencias de magnitudes como pueden ser la temperatura o la presión. Su importancia reside en que los gradientes son la auténtica fuente de la vida, que tiene que luchar contra la tendencia de la naturaleza a reducir gradientes, es decir, contra la tendencia de la energía a disiparse conforme a la segunda ley de la termodinámica (expresada según la a menudo citada entropía). Para los seres vivos, el equilibrio termodinámico equivale a la muerte, por lo que para com-

39

Esta conferencia no fue publicada en su momento; ha sido incluida en Lorenz (1995, 185-188).

prender la vida es imperativo entender la termodinámica del no equilibrio y no limitarse a la del equilibrio, que dominó gran parte de los siglos XIX y XX. La complejidad de la vida –y de otros sistemas existentes en la naturaleza– es una derivación natural de la tendencia a reducción de gradientes: allí donde las circunstancias lo permiten, surgen organizaciones cíclicas para disipar entropía en forma de calor. Puede incluso argumentarse –es una nueva forma, poco darwiniana, de entender la evolución– que puesto que el acceso a los gradientes se mejora mediante el perfeccionamiento de la percepción, el incremento de la inteligencia es una tendencia evolutiva que promueve selectivamente la prosperidad de aquellos que explotan recursos menguantes sin agotarlos. Esta rama de la física (y de la química) experimentó un gran desarrollo en la segunda mitad del siglo XX, y por ello constituye un magnífico ejemplo de otros avances que han tenido lugar a lo largo de ese periodo en la física y que, como decía, tal vez debería haber tratado aquí, aunque sean en cierto sentido de un carácter «menos fundamental». Pero ya me he extendido demasiado y es hora de poner punto final.

Bibliografía

- Aczel, A. D. *Entrelazamiento*. Barcelona: Crítica, 2004 (edición original en inglés de 2002).
- Allen, J. y D. Misener. «Flow of liquid helium II». *Nature* 141 (1938): 75.
- Aspect, A., J. Dalibard y G. Roger. «Experimental test of Bell's inequalities using time-varying analyzers». *Physical Review Letters* 49 (1982): 1.804-1.807.
- Bardeen, J. y W. Brattain. «The transistor, a semiconductor triode». *Physical Review* 74 (1948): 230-231 (L).
- «Physical principles involved in transistor action». *Physical Review* 75 (1949): 1.208-1.225.
- L. N. Cooper y J. R. Schrieffer. «Microscopic theory of superconductivity». *Physical Review* 106 (1957): 162-164 (L).
- Basov, N. G. y A. M. Prokhorov. «3-level gas oscillator». *Eksperim. i Teor. Fiz. (JETP)* 27 (1954): 431.
- Bednorz, J. G. y K. A. Müller. «Possible high T_c superconductivity in the Ba-La-Cu-O system». *Zeitschrift für Physik B- Condensed Matter* 64 (1986): 189-193.
- Bell, J. S. «On the Einstein-Podolsky-Rosen paradox». *Physics* 1 (1964): 195-200.
- «On the problem of hidden variables in quantum mechanics». *Reviews of Modern Physics* 38 (1966): 447-452..
- Bethe, H. «Energy production on stars». *Physical Review* 55 (1939a): 103.
- «Energy production in stars», *Physical Review* 55 (1939b): 434-456.
- y C. L. Critchfield. «The formation of deuterons by proton combination», *Physics Review* 54 (1938): 248-254.
- Bondi, H. y Th. Gold. «The steady-state theory of the expanding universe». *Monthly Notices of the Royal Astronomical Society* 108 (1948): 252-270.
- Born, M. «Zur Quantenmechanik der Stossvorgänge. (Vorläufige Mitteilung)». *Zeitschrift für Physik* 37 (1926): 863-867.
- Bose, S. «Plancks Gesetz und Lichtquantenhypothese». *Zeitschrift für Physik* 26 (1924): 178-181.
- Chandrasekhar, S. «The maximum mass of ideal white dwarfs». *Astrophysical Journal* 74 (1932): 81-82.
- Chu, P. C. W. «Superconductivity above 90 K». *Proceedings of the National Academy of Sciences* 84 (1987): 4.681-4.682.

- Clauser, J. F., M. A. Horne, A. Shimony y R. A. Holt. «Proposed experiment to test local hidden-variable theories». *Physical Review Letters* 23, (1969): 880-884.
- Collins, R. J., D. F. Nelson, A. L. Schawlow, W. Bond, C. G. B. Garret y W. Kaiser. «Coherence, narrowing, directionality, and relaxation oscillations in the light emission from ruby». *Physical Review Letters* 5 (1960): 303-305.
- Cooper, L. N. «Bound electron pairs in a degenerate Fermi gas». *Physical Review* 104 (1956): 1.189-1.190 (L).
- Cornell, E. A. y C. E. Wiemann. «El condensado de Bose-Einstein». *Investigación y Ciencia. Temas* 31 (primer trimestre 2003): 82-87.
- Einstein, A. «Zur Elektrodynamik bewegter Körper». *Annalen der Physik* 17 (1905a): 891-921.
- «Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt». *Annalen der Physik* 17 (1905b): 132-148.
- «Die Feldgleichungen der Gravitation». *Preussische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1915): 844-847.
- «Strahlungs -Emission und -Absorption nach der Quantentheorie». *Deutsche Physikalische Gesellschaft. Verhandlungen* 18 (1916a): 318-323.
- «Zur Quantentheorie der Strahlung». *Physikalische Gesellschaft Zürich. Mitteilungen* 18 (1916b): 47-62. También publicado en *Physikalische Zeitschrift* 18 (1917): 121-128.
- «Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie», *Preussische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1917): 142-152.
- «Quantentheorie des einatomigen idealen Gases». *Preussische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1924): 261-267.
- «Quantentheorie des einatomigen idealen gases. 2. Abhandlung». *Preussische Akademie der Wissenschaften (Berlin). Sitzungsberichte* (1925): 3-14.
- «Lens-like action of a star by the deviation of light in the gravitational field». *Science* 84 (1936): 506-507.
- B. Podolsky y N. Rosen. «Can quantum mechanical description of physical reality be considered complete?». *Physical Review* 47 (1935): 777-780.
- Fermi, E. «Tentativo di una teoria dei raggi». *Il Nuovo Cimento* 11 (1934a): 1-19.
- «Versuch einer Theorie der β -Strahlen. I». *Zeitschrift für Physik* 88 (1934b): 161-177.
- Feynman, R. P. «Space-time approach to quantum electrodynamics». *Physical Review* 76 (1949): 769-789.
- y M. Gell-Mann. «Theory of the Fermi interaction». *Physical Review* 109 (1958): 193-198.
- Fukuda, H., Y. Miyamoto y S. Tomonaga. «A self-consistent subtraction method in the quantum field theory. II». *Progress in Theoretical Physics* 4 (1939): 47-59.
- Gell-Mann, M. «A schematic model of baryons and mesons». *Physic Letters* 8 (1964): 214-215.
- *El quark y el jaguar*. Barcelona: Tusquets, 1995 (edición original en inglés de 1994).
- Georgi, H. y S. L. Glashow. «Unity of all elementary particle forces». *Physical Review Letters* (1974): 438.
- Glashow, S. L. «Partial-symmetries of weak interactions». *Nuclear Physics* 22 (1960): 579-588.
- Gold, T. «Rotating neutron stars as the origin of the pulsating radio sources». *Nature* 218 (1968): 731-732.
- Gordon, J., P. Herbert, J. Zeiger y Ch. H. Townes. «Molecular microwave oscillator and new hyperfine structure in the microwave spectrum of NH_3 ». *Physical Review* 95 (1954): 282-284 (L).
- Greene, B. *El universo elegante*. Barcelona: Crítica / Planeta, 2001 (edición original en inglés de 1999).
- Guth, A. H. «Inflationary universe: a possible solution to the horizon and flatness problem». *Physical Review D* 23 (1981): 347-356.
- Hawking, S. W. «Occurrence of singularities in open universes». *Physical Review Letters* 15 (1965): 689
- «Occurrence of singularities in cosmology». *Proceedings Royal Society A* 294 (1966a): 511-521.
- «Occurrence of singularities in cosmology». *Proceedings Royal Society A* 295 (1966b): 490-493.
- «Particle creation by black holes», en Isham, Penrose y Sciama, eds. (1975): 219-267.
- y R. Penrose. «The singularities of gravitational collapse and cosmology». *Proceedings of the Royal Society of London A* 314 (1969): 529-548.
- Heisenberg, W. «Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen». *Zeitschrift für Physik* 33 (1925): 879-893.
- «Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik». *Zeitschrift für Physik* 43 (1927): 172-198.
- Hewish, A., S. J. Bell, J. D. H. Pilkington, P. F. Scott y R. A. Collins. «Observation of a rapidly pulsating radio source». *Nature* 217 (1968): 709-713.
- Hoddeson, L., L. Brown, M. Riordan y M. Dresden, (eds). *The Rise of the Standard Model*. Cambridge: Cambridge University Press, 1997.
- Hoyle, F. «A new model for the expanding universe». *Monthly Notices of the Royal Astronomical Society* 108 (1948): 372-382.
- Hubble, E. «A relation between distance and radial velocity among extra-galactic nebulae». *Proceedings of the National Academy of Sciences of the U.S.A.* 15, (1929): 168-173.
- y M. L. Humanson. «The velocity-distance relation among extra-galactic nebulae». *Astrophysical Journal* 74 (1931): 43-80.
- Hulse, R. A. y J. H. Taylor. «Discovery of a pulsar in a binary system». *Astrophysical Journal* 195 (1975): L51-L53.
- Isham, Ch. J., R. Penrose y D. W. Sciama, (eds). *Quantum Gravity. An Oxford Symposium*. Oxford: Oxford University Press, 1975.
- Kamerlingh Onnes, H. «Further experiments with liquid helium. C. On the change of electric resistance of pure metals at very low temperature. IV. The resistance of pure mercury at helium temperatures». *Communications from the Physical Laboratory at the University of Leiden* N°. 120a (1911): 3-15.
- Kapitza, P. «Viscosity of liquid helium above the λ -point». *Nature* 141 (1938): 74.
- Landau, L. «On the theory of stars». *Phys. Z. Sowjetunion* 1 (1932): 285-287.
- Lemaître, G. «Un univers homogène de masse constante et de rayon croissant, rendant compte de la vitesse radiale des nébuleuses extra-galactiques». *Annales de la Société Scientifique de Bruxelles* 47 (1927): 49-59.
- Linde, A. D. «A new inflationary universe scenario: a possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problem». *Physics Letters* 108 D (1982): 383-393.
- London, F. «The λ -point phenomenon of liquid helium and the Bose-Einstein degeneracy». *Nature* 141 (1938): 643-644.
- Lorenz, E. N. «Deterministic non-periodic flows». *Journal of Atmospheric Science* 20 (1963): 130-141.
- *La esencia del caos*. Madrid: Debate, 1995 (edición original en inglés de 1993).

- Maiman, Th. H. «Stimulated optical radiation in ruby». *Nature* 187 (1960): 493-494.
- Mather, J. C. et al. «A preliminary measurement of the cosmic microwave background spectrum by the Cosmic Background Explorer (COBE) satellite». *Astrophysical Journal* 354 (1990): L-37-L40.
- Mayor, M. y D. Queloz. «A Jupiter-mass companion to a solar-type star». *Nature* 378 (1995): 355-359.
- Müller, K. A. y J. G. Bednorz. «The discovery of a class of high-temperature superconductivity». *Science* 237 (1987): 1133-1139.
- Oppenheimer, J. R. y H. Snyder. «On continued gravitational contraction». *Physical Review* 56 (1939): 455-459.
- y G. Volkov. «On massive neutron cores». *Physical Review* 55 (1939): 374-381.
- Penrose, R. «Gravitational collapse and space-time singularities». *Physical Review Letters* 14 (1965): 57-59.
- Penzias, A. A. y R. W. Wilson. «A measurement of excess antenna temperature at 4080 Mc/s». *Astrophysical Journal* 142 (1965): 414-419.
- Perlmutter, S. et al. «Discovery of a supernova explosion at half the age of the universe and its cosmological implications». *Nature* 391 (1998): 51-54.
- Planck, M. «Zur Theorie des Gesetzes der Energieverteilung im Normalspektrum». *Verhandlungen der Deutschen Physikalischen Gesellschaft* 2 (1900): 237-243.
- Ryle, M. «Radio stars and their cosmological significance». *The Observatory* 75 (1955): 137-147.
- Salam, A. «Weak and electromagnetic interactions», en Svartholm, ed. (1968).
- Schawlow, A. y Ch. H. Townes. «Infrared and optical masers». *Physical Review* 112 (1958): 324-327.
- «A medium in which a condition of population inversion exists». U.S. Patent 2.929.922, 22 de marzo de 1960.
- Schödinger, E. «Quantisierung als Eigenwertproblem. (Erste Mitteilung)». *Annalen der Physik* 79 (1926): 361-376.
- Schmidt, B. et al. «High-Z supernova search: Measuring cosmic deceleration and global curvature of the universe using type Ia supernova». *Astrophysical Journal* 507 (1998): 46-63.
- Schweber, S. «A historical perspective on the rise of the standard model». En Hoddeson, Brown, Riordan y Dresden, eds. (1997): 645-684.
- Schwinger, J. S. «On radiative corrections to electron scattering». *Physical Review* 75 (1949): 898-899 (L).
- Shockley, W. «Density of surface states on silicon deduced from contact potential measurements». *Physical Review* 72 (1947): 345.
- «Modulation of conductance of thin films of semiconductors by surface changes». *Physical Review* 74 (1948): 232-233.
- Smolin, L. *Las dudas de la física en el siglo XXI*. Barcelona: Crítica, 2007 (edición original en inglés de 2006).
- Smoot, G. et al. «Structure of the COBE differential microwave radiometer first year maps». *Astrophysical Journal* 396 (1992): L1-L5.
- y K. Davison. *Arrugas en el tiempo*. Barcelona: Círculo de Lectores, 1994 (edición original en inglés de 1993).
- Sudarshan, E., C. George y R. E. Marshak. «The nature of the four-fermion interaction». *Padua Conference on Mesons and Recently Discovered Particles*. Padua, 1957: V14-24.
- «Chirality invariance and the universal Fermi interaction». *Physical Review* 109 (1958): 1.860-1.862.
- Svartholm, N., ed. *Elementary Particle Theory: Relativistic Groups and Analyticity*. Estocolmo: Almqvist and Wilksell, 1968.
- Taylor, J. H, L. A. Fowler y P. M. McCulloch. «Measurements of general relativistic effects in the binary pulsar PSR1913+16». *Nature* 277 (1979): 437-440.
- Tisza, L. «Transport phenomena in helium II». *Nature* 141 (1938): 913.
- Wagoner, R. V., W. A. Fowler y F. Hoyle. «On the synthesis of elements at very high temperatures». *Astrophysical Journal* 148 (1967): 3-49.
- Walsh, D., R. F. Carswell y R. J. Weyman. «0957+561 {A}, {B}: twin quasistellar objects or gravitational lens?». *Nature* 279 (1979): 381.
- Weinberg, S. «A model of leptons». *Physics Review Letters* 19 (1967): 1.264-1.266.
- *The First Three Minutes: A Modern View of the Origin of the Universe*. Nueva York: Basic Books, 1979.
- Weizsäcker, C. F. von. «Über elementumwandlungen im inner der sterne, II». *Physikalische Zeitschrift* 39 (1938): 633-646.
- Wheeler, J. A. y K. Ford. *Geons, Black Holes and Quantum Foam*. Nueva York: Norton, 1998.
- Witten, E. «String theory dynamics in various dimensions». *Nuclear Physics B* 443 (1995): 85-126.
- Wolszczan, A. y D. Frail. «A planetary system around the millisecond pulsar PSR1257+12». *Nature* 355 (1992): 145-147.
- Yoshimura, M. «Unified gauge theories and the baryon number of the universe». *Physical Review Letters* 41 (1978): 281-284.
- Zabusky, N. J. y M. D. Kruskal. «Interaction of "solitons" in a collisionless plasma and the recurrence of initial states». *Physical Review Letters* 15 (1965): 240-243.
- Zweig, G. «An SU(3) model for strong interaction symmetry and its breaking». *CERN Report No. 8181/Th 401* (enero 1964) y No. 8419/Th 412 (febrero 1964).

el arte de lo invisible. logros, beneficios sociales y desafíos de la nanotecnología

SANDIP TIWARI Y ROBERT MCGINN

1

En el Este, la Universidad de Nalanda existió desde el siglo v al xii. La universidad oriental más antigua que pervive en la actualidad es la de Al-Karaouine, en Marruecos, fundada en 859, y de las occidentales, la de Bolonia, de 1088. Las primeras universidades surgieron a partir de instituciones religiosas y fueron ganando independencia conforme declinaba el poder de la jerarquía eclesiástica.

2

J. Craig Venter, un importante contribuyente a la secuenciación de los genes humanos en la compañía Celera Genomics dirige hoy el J. Craig Venter Institute, una entidad autofinanciada dedicada a la investigación, cuyo éxito más reciente ha sido el progreso hacia la construcción de la célula artificial. Leroy Hood, que estuvo entre los pioneros del diseño de herramientas para el diagnóstico molecular, dirige el Institute for Systems Biology, una institución independiente. Estos ejemplos no difieren gran cosa de los de Thomas Edison, Graham Bell o Nikola Tesla a finales del siglo xix.

La historia de la ciencia y la ingeniería como fuerzas sociales de importancia es relativamente breve. La mayoría situarían su origen en la revolución copernicana del siglo xvi, es decir, que su existencia equivaldría a menos de la cuarta parte del tiempo que el hombre lleva sobre la Tierra. Con la llegada del progreso científico —y su empleo de herramientas agnósticas de matemáticas, cuestionamiento, postulación, teorización, predicción, verificación, creer en teorías lo suficiente como para seguir adelante en las investigaciones pero al tiempo dudando lo bastante como para detectar errores y fallos— llegó también el enfoque moderno del aprendizaje y la invención. Superar los dogmas, incluso frente a observaciones contradictorias, puede constituir un desafío para la sociedad y así será siempre; la comodidad que supone hacer «el trabajo de siempre» no es algo a desdeñar. Todo esto es cierto también en la empresa científica. Pero las ciencias físicas y de la naturaleza, la ingeniería y las matemáticas en tanto que áreas de trabajo son de las pocas empresas en las que las revoluciones pueden ocurrir con relativa facilidad. La teoría de la relatividad de Einstein —el «absoluto» de la velocidad de la luz y la gravedad como deformación del espacio-tiempo, la mecánica cuántica como una concepción completamente nueva de la mecánica para describir la realidad basada en enfoques probabilísticos—; también la comprensión filosófica de la realidad como resultado de la observación; el teorema de Gödel sobre los límites de la «demostrabilidad» dentro de un sistema

axiomático; la descodificación genómica del origen de la vida y la comprensión del metabolismo, la replicación y la reproducción, son todas ideas que fueron rápidamente adoptadas por la comunidad técnica una vez superaron el examen del enfoque científico.

La búsqueda científica de verdades y el impulso de ponerlas en práctica no conocen fronteras nacionales y se adaptan a las condiciones de cada época. Entre los progenitores de los albores de la civilización científica, Copérnico fue un hombre de Iglesia y polaco; Bruno, que pagó con su vida la defensa de sus creencias frente al dogma, fue teólogo de Italia; Tycho de Brahe, el matemático imperial, era de Dinamarca, y Johannes Kepler, alemán, y Galileo Galilei fueron profesores en Italia. No todos los pioneros de la ciencia fueron profesores, aunque las universidades como instituciones dedicadas a la enseñanza¹ existen desde hace mucho tiempo. En el siglo pasado Albert Einstein comenzó como abogado de patentes, y las contribuciones científicas de Neils Bohr surgieron de su estilo de vida nómada, que le hacía alejarse de la institución estatal donde trabajaba, un poco a la manera de Copérnico y Kepler. En la actualidad, cuando el poder y el impacto económico de la ciencia y la ingeniería han aumentado de forma drástica, se han fundado numerosas instituciones dedicadas a la investigación impulsadas por los científicos mismos² que se autofinancian o bien están financiadas por filántropos y otra suerte de mecenas modernos: inversores de capital riesgo y fundadores de pequeñas com-

pañías que se han enriquecido inmensamente gracias a aplicaciones de ciencia y de ingeniería. Las universidades, como en el pasado, desempeñan su papel, pero no son los únicos agentes de progreso. Los laboratorios fundados por el Estado, los independientes y los industriales, en especial los dedicados a las ciencias biológicas, colaboran también en la investigación científica y en sus aplicaciones.

Una revolución científica se origina a partir de individuos poseedores de una increíble voluntad y fortaleza interior, personas capaces de generar una poderosa fuerza centrípeta a partir de un único concepto como principio organizador universal.³ El progreso científico, periodo de consolidación, se produce porque hay individuos que persiguen varias conclusiones de forma centrífuga empleando una gran variedad de recursos para aprovechar las conexiones centradas en el principio organizador de un mundo lleno de complejidad. El progreso en la ciencia y la ingeniería depende tanto del descubrimiento central como de su posterior articulación. La creación de Mendeleev de la tabla periódica antes de que se conocieran las partículas y los átomos; la formulación de Darwin del principio de la evolución sin ninguna clase de conocimiento molecular, genético y orgánico, y el desarrollo de Heisenberg de la mecánica cuántica que reemplazó el determinismo newtoniano son ejemplos de superación de dogmas y creación de principios científicos nuevos.

Darse cuenta de que la química, la biología y la física tal como las conocemos y las empleamos hoy en ingeniería y medicina prácticamente no existieron hasta hace sólo un siglo y medio es toda una lección de humildad. A partir del descubrimiento y comprensión de los elementos químicos pronto desarrollamos la capacidad de fabricar amoníaco, y a partir de ahí fertilizantes para la agricultura que hacen posible la existencia de casi 7.000 millones de personas sobre la Tierra. Las interacciones genéticas y la progresiva comprensión de las mutaciones constituyen enfoques fundamentales en la lucha contra las enfermedades y harán posible que las personas vivan mejor y más tiempo. La informática y las comunicaciones, que dependen directamente de la electrónica, desarrollan los principios de su *hardware* a partir de la mecánica cuántica y de la teoría de la información. Somos muy afortunados de vivir en una época de descubrimientos y asistir a la aventura de la aplicación práctica de todos ellos.

Los avances centrífugos también dependen de la disponibilidad de herramientas: instrumentos de observación y creación. Cuanto más pequeña es la herramienta, mayores son las probabilidades de que sea personalizada o, lo que es lo mismo, individualizada y humanizada; es decir, de más fácil manejo. Debido a esta cualidad las herramientas son empleadas por numerosos científicos e ingenieros, que estimulan así su creatividad, lo que a su vez repercutirá en gran parte de la sociedad. El molino de agua evolucionó hasta convertirse en la máquina de vapor, y de ahí al motor eléctrico, de combustión, etcétera. Cada uno surgió

bajo muchas formas. El motor de combustión, con distintos formatos, propulsa al avión, el coche y la motocicleta. El motor eléctrico hace funcionar el tren, el aire acondicionado e incluso el disco duro de un ordenador portátil. Hoy sabemos que existe un motor de energía molecular, llamado ATP sintasa, que convierte energía química en movimiento mecánico dentro de nuestro organismo. ¿Quién sabe qué puertas abrirán este descubrimiento y sus reproducciones sintéticas en el laboratorio? Pero el propósito común que ha impulsado este proceso de miniaturización es el de encontrar aplicaciones que son útiles a los seres humanos. Los procedimientos médicos han cambiado drásticamente gracias a las herramientas endoscópicas; en la mayoría de los casos los ingresos prolongados en hospitales se han eliminado, el teléfono móvil y otros instrumentos de comunicación están por todas partes, incluso en las regiones más pobres del mundo, supuestamente con grandes beneficios derivados de un mejor intercambio de información. Los programas de *software* para escribir, dibujar y visualizar en nuestros pequeños ordenadores son innumerales. En todos estos casos la miniaturización y la personalización han tenido un impacto espectacular.

El progreso tecnológico tiene también, por supuesto, su lado oscuro en función de las innovaciones mismas y de la forma en que son difundidas y utilizadas. Los fertilizantes, los ordenadores y los motores de combustión consumen grandes cantidades de energía,⁴ son agentes contaminantes y han alterado el equilibrio del planeta. La Revolución Industrial en Europa redujo de forma drástica la esperanza media de vida. Gran parte de la energía consumida en el mundo hoy tardó miles de millones de años en acumularse en nuestro planeta, haciendo posible la existencia de 7.000 millones de humanos, en lugar de, tal vez, 1.000 millones, pero a cambio afectando el clima global. El impacto de las herramientas personalizadas es de efecto multiplicador. Los coches son un buen ejemplo de ello, pero también lo es el teléfono móvil. Cada nueva creación y las nuevas maneras en las que interactúan la sociedad en su conjunto y los individuos entre sí generan una nueva línea divisoria entre tener y no tener, entre aquellos que se adaptan y los que no, entre quienes aprenden y se benefician económica y socialmente de las herramientas y aquellos que no lo hacen. De manera que, mientras la media puede ser elevada, suelen aparecer desigualdades. Debido a que la tecnología a menudo facilita las tareas manuales, aquellos que se encuentran en el peldaño inferior de la escalera económica son más susceptibles de salir perjudicados con las nuevas tecnologías.

En este contexto social analizaremos ahora la promesa y los desafíos de la nueva y floreciente era de la especialización técnica: la ciencia, la ingeniería y la tecnología a nanoescala, a menudo resumidos con el término «nanotecnología». Básicamente la nanotecnología es un fenómeno relativo al tamaño, a la dimensión. Al igual que la biología, que abarca una gran variedad de ciencias de la

³ Para citar a Arthur Koestler, «cuanto más original es un descubrimiento, más obvio resulta a posteriori».

⁴ Se considera que las industrias de fertilizantes y de la información suponen casi un 10% del consumo total de energía, y el motor de combustión aún más.

vida, la nanotecnología afecta a aquellas áreas de la ciencia, la ingeniería y la tecnología donde interviene la escala. Tal vez en el futuro decidamos llamarla «nanología» para reflejar así este amplio radio de influencia, en lugar de hablar de nanotecnología, nanociencia y nanoingeniería.

Si tomamos cualquier material a granel que podamos ver con los ojos, ya se trate de material duro o blando, orgánico o inorgánico, y lo hacemos más pequeño, sigue conservando algunas propiedades. Un trozo grande de diamante, de hierro o de caucho posee las mismas propiedades que uno pequeño, y estos materiales y su reproducibilidad nos serán más útiles cuando conservan sus propiedades después de reducidas sus dimensiones. Los puentes pueden ser grandes y pequeños, un solo carril para coches sobre un riachuelo o lo suficientemente amplios para permitir que un tren cruce el océano. El plástico se emplea en los coches y también en pequeños relojes de pulsera. Por otra parte, si vamos al extremo contrario y reducimos el tamaño de un material al máximo, es decir, a escala atómica o molecular, sus propiedades serán completamente distintas. Un átomo o molécula tiene propiedades que resultan de las interacciones de mecánica cuántica y que conducen a su existencia como unidad estable. El carbón a escala atómica forma diamantes, pero también grafito, y además es el principal componente del hollín, resultado de combustión ineficiente. Todas estas formas de carbón poseen diferentes propiedades, de manera que la transición de la escala atómica a la escala a granel implica una transformación en las propiedades del material, que se mide en nanómetros. Las propiedades de los materiales y sus interacciones físicas y químicas surgen de las fuerzas de la naturaleza: el enlace atómico o molecular es resultado de la mecánica cuántica y de fuerzas electromagnéticas. Por medio de las interacciones de y entre átomos, moléculas, electrones —portadores de corriente eléctrica—, y fotones —portadores de luz— los procesos químicos y biológicos experimentan una transformación radical de sus propiedades a nanoescala. Ello resulta en que el átomo y la molécula se ensamblan por un lado y los materiales a granel por otro. La razón es que las fuerzas que están en el centro de estas interacciones, las que resultan de las propiedades características, son básicamente de naturaleza nanométrica.

Y los cambios a nanoescala no son simplemente más pequeños; pueden ser tan radicales como la aparición de nuevas propiedades a las que antes no teníamos acceso ni a macro ni a microescala.

El tunelado de la mecánica cuántica es un fenómeno que se ha empleado con éxito en la última década: en memorias semiconductoras que no pierden sus datos y no tienen parte móvil, como las que emplean una cámara fotográfica, un teléfono móvil y el *thumb drive* o unidad de almacenamiento en miniatura. En estos mecanismos los electrones circulan por una región de aislamiento magnético a bajo voltaje (tunelado). Ello sucede gracias

a la naturaleza ondulante del electrón y la capacidad de la onda de penetrar pequeñas distancias —distancias de nanoescala— que generan un aislamiento magnético. Las propiedades básicas tales como temperatura de fundición, magnetización, capacidad de corriente eléctrica, etcétera, pueden alterarse sin que cambie la composición química del material, debido precisamente a este comportamiento ondulante y a las interacciones que se producen a nanoescala, una propiedad que los sopladores de vidrio de la Edad Media empleaban en la fabricación de vidrio policromado. El vidrio policromado a menudo emplea nanopartículas de oro y plata que crean el rojo, el azul, el marrón y otros colores resaltando la diseminación de un color particular dependiendo del tamaño de las partículas. Los electrones presentes en nanopartículas de oro y plata interactúan con los fotones de luz creando el color. Los científicos describen esta interacción colectiva entre electrones de plasma y fotones mediante una partícula a la que llaman «plasmón». Los sopladores de vidrio desarrollaron sin saberlo la tecnología que permite precipitar estos plasmones de forma controlada a ese tamaño. La luz que transportan las fibras ópticas y que hace posible la rápida transmisión de datos se fabrica empleando láseres que constituyen fuentes de luz altamente eficientes por medio de interacciones entre electrones y fotones que surgen a nanoescala en pozos cuánticos artificialmente creados. Incluso la fibra óptica emplea confinamiento de luz a nanoescala para desplazarla con las mínimas pérdidas posibles a través de largas distancias.

Las reacciones químicas resultan de interacciones entre átomos y moléculas en estado neutro o activado o excitado. Las especies resultantes de esta reacción necesitan acercarse y disponer de vías favorables desde el punto de vista energético para que las reacciones sean efectivas. La catálisis es central a la hora de hacer posible esto: un catalizador, aunque permanezca químicamente inalterado, proporciona una vía de baja energía a los índices crecientes de reacción. Para ello crea una superficie en la que se juntan las moléculas y reaccionan unas con las otras en condiciones energéticamente favorables, dejando el catalizador inalterado una vez finalizadas las reacciones. Cuando descendemos al nivel de las pequeñas dimensiones, aumenta la relación superficie/volumen, un efecto de la nanoescala. Un mero reforzamiento no lineal de esta sencilla propiedad hace la catálisis enormemente efectiva. El proceso Haber-Bösch de producción de amoníaco, un ingrediente clave en la fabricación de fertilizantes, emplea la catálisis en muchas de sus etapas. El hidrógeno se obtiene del metano presente en el gas natural empleando óxido de níquel. El amoníaco se forma a partir de nitrógeno y de este hidrógeno empleando hierro, derivado de la magnetita, con una eficiencia de conversión última del 99%, es decir, casi perfecta.

La magnetita, una forma de óxido de hierro, es un material cuyas propiedades nanométricas ha empleado la naturaleza durante siglos, desde antes incluso que existie-

ran los sopladores de vidrio. La magnetita es magnética. Puesto que está hecha de una colección de nanocristales ordenados en forma de cadenas y que por tanto constituyen un magneto altamente sensible, dota a los organismos de una propiedad llamada magnetotaxis, que es la capacidad de responder a los campos magnéticos de la Tierra. Así, la *Magnetospirillum magnetotacticum*, una bacteria que se encuentra en estanques y que fue aislada por vez primera en 1975 junto con muchas otras, es magnetotáctica porque en pequeña escala y formando un conjunto, un organismo primitivo puede identificar desviaciones de los campos magnéticos terrestres. Muchas especies animales emplean la información magnética para orientarse, incluidas la paloma, la tortuga boba y la langosta espinosa. En el proceso evolutivo la naturaleza desarrolló maneras mediante las que los nanocristales inorgánicos podían agruparse en sistemas mayoritariamente orgánicos, algo que aún estamos aprendiendo a hacer de forma controlada en el laboratorio. Otro ejemplo interesante de escala nanométrica en la naturaleza es el color iridiscente de algunas mariposas y plumas de pavo real. Son efectos ópticos a nanoescala resultantes de las estructuras tridimensionales que crea la naturaleza, y su recreación en laboratorio se encuentra todavía en fase temprana. Los fenómenos biológicos tienden a ser inmensamente complejos, al ser resultado de una combinación aleatoria de hechos y de un gran número de interacciones que se producen entre gran número de entidades bajo la influencia de fuerzas locales. Estos fenómenos son sensibles a las condiciones iniciales y a mínimas perturbaciones, tienen un gran número de componentes en continua interacción y a menudo también un gran número de vías por las que puede evolucionar el sistema. Si un ser humano recibe una cantidad insuficiente de energía, es decir, si no ha comido lo suficiente, su cuerpo sabe cómo ralentizar el metabolismo. A diferencia de mucho de lo que hacemos en los campos de la física, la química y la ingeniería, esto es algo mucho más complejo e implica una variedad de interacciones a diferentes escalas. Que esos organismos simples y complejos hayan encontrado la manera de hacer que dominios magnéticos a nanoescala de un solo cristal adquieran estas propiedades es un tributo a los recursos de la naturaleza y a la inteligencia de la biología, características que la especie humana descubre con regularidad.

Las últimas décadas prepararon el terreno para el desarrollo de la ciencia y la ingeniería de la materia condensada, donde el empleo de herramientas pequeñas y personalizadas se difundió y la capacidad de controlar y observar a nanoescala se volvió accesible para un gran número de personas. Estas herramientas nos permiten ensamblar, manipular, controlar, sondear, fotografiar y observar una miríada de propiedades a nanoescala. De todas las herramientas, las que más publicidad han recibido son el microscopio tunelador escáner y la microscopía de fuerza atómica. Pero igualmente importantes han sido

las herramientas de fabricación que nos permiten definir y ensamblar a escala nanométrica, las nuevas técnicas para visualización, los instrumentos que permiten el autoensamblaje de monocapas en superficies, las herramientas que hacen posible sintetizar y, en general, las que nos permiten hacer todo esto en serie, rápidamente y a bajo coste. Ahora tenemos la capacidad de sintetizar átomo a átomo y también de esculpir los materiales hasta descender el nivel atómico. Podemos testar fenómenos que se producen a nanoescala a través de una amplia variedad de herramientas, que a su vez hacen posible una variedad de enfoques. Y debido a que las propiedades cambian radicalmente cuando se desciende al nivel de las unidades más pequeñas, podemos actuar sobre ellas mediante técnicas de ensamblado y escultura. Esto, a su vez, ha hecho posible que una amplia comunidad de individuos pueda descender al mundo de la nanoescala. La nanoescala es una dimensión, no una disciplina, y sus propiedades se ponen de manifiesto en y están relacionadas con todas las disciplinas. El resultado de esto, a través de la participación a gran escala de la comunidad y la amplitud de disciplinas, ha sido que en las interfaces ha surgido un nuevo y gran campo de trabajo. La ingeniería, las ciencias físicas y de la vida han convergido como nunca antes lo habían hecho. Y ello ha conducido a niveles de progreso y utilidad que hasta hace una década resultaban inimaginables.

Unos pocos ejemplos de esta variedad, que reside en el corazón de la existencia humana, ilustrarán este punto. Examinemos algunos de los desafíos a los que en la actualidad se enfrenta el mundo. Los más importantes tienen que ver con la sostenibilidad, con que una comunidad compleja y cada vez más amplia de personas y de millones de otras especies puedan vivir de forma sostenible, a saber, en equilibrio las unas con las otras y con el mundo natural. La energía, la salud, la igualdad, la pobreza, la educación y la conservación son asuntos que nos vienen inmediatamente a la cabeza como pasos previos para la sostenibilidad. Pronto surgen también determinadas preguntas: ¿podemos reducir el consumo de energía en transportes, en iluminación, producción de alimentos y otras facetas de la vida diaria mediante una recreación de nuestro entorno (calefacción, refrigeración y estética) y de las comunicaciones (en el intercambio de información, en informática y en todos los instrumentos móviles)? ¿Podemos ayudar a solucionar los problemas de escasez de agua produciendo agua limpia, eliminando impurezas derivadas de los metales pesados tales como el arsénico y reduciendo el consumo de agua? ¿Podemos mejorar la productividad agrícola desarrollando plantas de consumo saludable que sean más resistentes a las enfermedades y consuman menos energía y menos agua? ¿Podemos diseñar un secuestro más eficiente del carbono empleando enfoques físicos y biológicos? ¿Podemos mejorar la gestión de los recursos forestales consumiendo menos papel e introduciendo técnicas de producción de papel más efi-

cientes? ¿Podemos mejorar la asistencia sanitaria haciendo posible diagnósticos más tempranos y menos costosos, detectar la contaminación, curar enfermedades, mejorar el tratamiento o ralentizar el avance de las enfermedades degenerativas y atacar las más perniciosas como la malaria y el cáncer? La nanotecnología encierra la promesa de una posible respuesta a todas estas preguntas.

La fortaleza de los materiales y las propiedades de superficie de éstos se emplean en todas partes. Los polímeros, cuya síntesis a escala industrial se generalizó a mediados del siglo xx, son hoy prácticamente ubicuos. Hay quien argumentaría que los plásticos fueron la columna vertebral de la revolución industrial de China y la clave de la transformación de la vida cotidiana, con su empleo en la fabricación de juguetes infantiles hasta las bolsas de la compra y envoltorios variados. Los plásticos y los polímeros adquieren sus propiedades mediante interacciones de superficie de cadenas de hidrocarburos y ambos se han visto beneficiados de los nuevos inventos nanotecnológicos. Los nanotubos de carbono, basados en un fuerte enlace entre los átomos de carbono —una configuración distinta de la del diamante— proporcionan propiedades de interacción de superficie intrínseca; soportan fuerzas más poderosas que el acero de dimensiones similares.⁵ Si se transforman en hebras, de manera similar a como se hace con los polímeros, tendremos materiales de gran resistencia. Los nanotubos de carbono se están empezando a incorporar a la fabricación de plásticos para hacerlos más resistentes, por ejemplo en material deportivo como raquetas de tenis y palos de golf. Compuestos como el hormigón, la fibra de vidrio y el Kevlar son combinaciones de materiales que se vuelven más fuertes gracias a interacciones de superficie. El hormigón puede hacerse más ligero y conservar no obstante su resistencia mediante el uso de cenizas, las estructuras huecas de sílice y alúmina parecidas a los fullerenos (molécula de carbono C₆₀) presentes en las cenizas de plantas de energía de carbón. La resistencia del nanomaterial y la poderosa interfaz hacen posible que estos compuestos sean más fuertes que nunca. La superficie es clave en esta propiedad.

Hemos mencionado la catálisis y su importancia en los procesos de producción de amoníaco como uno de los grandes avances del siglo xx. Hoy los zeolitos desempeñan un papel similar. Son sólidos microporosos que se transforman en eficientes catalizadores a partir de óxidos de aluminio y silicio. Usados en millones de toneladas ayudan a craquear el petróleo para la obtención de gasolina e hidrocarburos, reduciendo el impacto ambiental del petróleo.

Parece probable que los avances en nanotecnología tengan efectos muy positivos en la producción y consumo de energía, en las comunicaciones y en la salud. Consideremos algunos de ellos.

Las células de combustible, la energía fotoeléctrica y la conversión de energía fotoeléctrica son ejemplos de que las mejoras eficientes relacionadas con la energía están

aumentando rápidamente gracias a nuevos materiales, las membranas ultra delgadas y los nuevos procesos de conversión. Las fuentes de luz obtenidas mediante semiconductores son altamente eficientes, con un factor de sostenibilidad 10 frente a la bombilla incandescente; además resultan más fiables y duraderas. Hoy podemos verlos en semáforos, pero pronto también en el alumbrado general, en cuanto se resuelvan cuestiones relativas al coste y a la preferencia de ciertos colores. También se están creando fuentes de luz a partir de materiales orgánicos, aunque en este caso el desafío de la fiabilidad es mayor. La generación de luz fotovoltaica también se está beneficiando de los avances a nanoescala. Un nuevo tipo de célula solar, llamada Grätzer, comienza a hacer su transición de los laboratorios a la fabricación industrial. Estas células utilizan titanio nanocristalino, colorantes y materiales orgánicos en el transporte de electrones para alcanzar un porcentaje de eficiencia en la conversión energética. El titanio es un material que se encuentra en la pintura, en el papel de lija y en muchos otros lugares donde interviene la fuerza. También absorbe protones de manera eficiente y por tanto se emplea en lociones solares. Las nuevas estructuras fotovoltaicas emplean procesos de baja energía en su fabricación, a diferencia de la mayoría de los silicios fotovoltaicos populares hoy día, reduciendo también en consecuencia el coste y la energía requerida para su fabricación. Las interacciones reforzadas de superficie pueden ayudar a controlar la contaminación. En regiones densamente pobladas como la llanura Indogangéctica, el descenso de la pluviosidad ha dado lugar a grandes concentraciones de arsénico originado por causas naturales en los pozos de agua potable. La mayor eficiencia de los procesos electroquímicos aplicados a la superficie hace posible depurar el arsénico con ayuda de nanopartículas de óxido de hierro.

La electrónica, la informática y las comunicaciones se han beneficiado enormemente de las propiedades de la nanoescala, en la cual las ondas de electrones y el material interactúan de múltiples maneras para producir interesantes propiedades. Consideremos, por ejemplo, el almacenamiento de datos. En la actualidad la humanidad crea cada día más datos que la cantidad total de éstos almacenada hace 20 años. El almacenamiento semiconductor no volátil se emplea en cámaras de fotos, teléfonos, minirreproductores de música y también para almacenar e intercambiar información. Esto funciona porque tiene lugar, a nanoescala, un fenómeno de mecánica cuántica llamado «tunelado». El inmenso volumen de datos que procesa Google y que las empresas almacenan es viable porque las unidades de disco magnéticas almacenan más en menos espacio, es decir, son más compactas y también cuestan menos. Esto es posible porque se han aprovechado la rotación (*spin*) del electrón y las interacciones de campo que se producen a nanoescala. Nuestra veloz infraestructura de comunicaciones depende de la transmisión óptica. Lo diodos láser de pequeño tamaño y los amplificadores y las fibras ópti-

5

En la actualidad hay una enfermedad que se está extendiendo entre la comunidad científica: la sobreabundancia de supuestos hallazgos que bordean la incredulidad. Esta enfermedad, que siempre ha existido, es hoy particularmente pernicioso porque la amplia variedad de disciplinas existentes hace difícil a muchos discernir los hallazgos verdaderos de los que no lo son. Tal vez se trate de un problema social y ético, provocado por la presión que reciben los científicos para que justifiquen sus investigaciones. Existe también una escuela de pensamiento que, con un exceso de optimismo y euforia, anima a los jóvenes a que se dediquen a la investigación científica e ingenieril, una tendencia que se ha visto favorecida por la llegada de Internet, el fenómeno de la inmediatez y la facilidad para crear imágenes visuales poco realistas mediante *software* personalizado. Los nanotubos de carbono para supuestos ascensores espaciales son uno de estos mitos (ver: «The space elevator: going down?», en *Nature* online, publicado el 22 de mayo de 2006 y disponible en <http://www.nature.com/news/2006/060522/full/news060522-1.html>). Existen otras teorías semejantes relativas al uso de moléculas y otros enfoques a la electrónica a escala atómica.

cas emplean confinamiento de portadores y fotones en dimensiones reducidas para obtener una mayor eficiencia en generación y transmisión de señales. Los mecanismos de menor tamaño también consumen menos, de manera que el índice de consumo energético por mecanismo ha disminuido con el tiempo. Sin embargo, la personalización de pequeñas herramientas, por ejemplo, ordenadores, también significa que ha aumentado su número de usuarios. De ahí que las cifras generales sigan siendo las mismas.

El empleo generalizado en electrónica de estos mecanismos de detección y control también ha sido determinante en cómo se aplica la nanotecnología a la ciencia. Uno de los principales desafíos en biociencia ha sido la comprensión minuciosa de los fenómenos bajo las condiciones químicas y físicas específicas que se dan en entornos reales. La invención de la reacción en cadena de la polimerasa (PCR por sus siglas en inglés) proporcionó una técnica para amplificar una muestra de ADN que fuera de interés, posibilitando hacer tantas copias de la misma como se precisaran para su estudio y análisis. En forma de nanoherramienta, el PCR hizo posible generar millones de copias de una hebra concreta de ADN y usarlas para manipulación genética. Las técnicas de microconfiguración (*microarray*), monoclonales y de prótesis fluorescentes han resultado ser igualmente provechosas. Gran parte de la investigación biológica, sin embargo, continúa dependiendo del análisis estadístico de datos, en los cuales se producen una gran cantidad de interacciones como las mencionadas y a continuación se extraen de ellas los modelos posibles para describir la especificidad. Las técnicas de la ciencia física tienden a descartar todos los fenómenos superfluos y a simplificar el sistema de forma que las propiedades relevantes puedan estudiarse con el debido rigor. Con la aparición de numerosas «nanotécnicas», es decir, técnicas capaces de descender a la escala más pequeña que existe, comienza a ser posible evitar la dependencia del análisis estadístico y estudiar las posibilidades de manera exhaustiva.

Hacer esto, no obstante, requiere disponer de ultrasensores. La nanotecnología permite fabricar una gran variedad de ellos. Una molécula fluorescente puede ser reemplazada por una nanopartícula ópticamente más activa sintonizada con una longitud de onda específica y ligada a una molécula cuya química se está estudiando. Se pueden usar las interacciones plasmónicas (de electrones de plasma y electromagnéticas) para localizar el calor durante el emparejamiento energético a escala nanométrica. Las vigas empleadas en construcción pueden reducirse hasta el punto de que sea posible detectar la sensibilidad al peso a escala de un solo átomo. Las nanoherramientas pueden empujarse para aislar y testar células en enlaces bi- y tridimensionales. Es posible también usar pinzas ópticas para atrapar nanopartículas, desplazarlas y, si se quiere, estudiar sus distintas reacciones así como las moléculas unidas a ellas. Por tanto hoy día

resulta concebible colocar nanopartículas y otras herramientas para que observen e interactúen *en el interior* de células y tejidos y observarlas y representarlas ópticamente a tiempo real, lo que permitiría descifrar el complejo funcionamiento interno de la célula. Es posible trabajar con estas herramientas en condiciones realistas debido a los grandes avances realizados en detección, representación óptica y control que la nanoescala ha hecho posibles.

Los científicos tienden a menospreciar lo que es posible lograr a corto plazo —digamos diez años— y también lo que puede hacerse a largo plazo: cincuenta años. Lo que resulta muy interesante de la nanotecnología es que, debido a que se basa en la escala nanométrica, tiene aplicaciones en muy diversas disciplinas. Nunca antes en la historia de la humanidad han generado los científicos información de tanta utilidad para los más variados ámbitos. La última década ha sido un buen comienzo, pero conforme se desarrollen las herramientas y la comprensión de la nanotecnología se descubrirán nuevas aplicaciones que surgirán de la intersección entre distintas disciplinas. El progreso deberá mantener su ritmo tanto en la ciencia física como en la ingeniería, de manera que la iluminación fotovoltaica, la informática de consumo eficiente, el almacenaje y obtención de información y las comunicaciones continúen avanzando. Puede argumentarse, con razón, que la química y la ciencia de los materiales deberían haberse centrado en la nanotecnología desde el principio; después de todo, la catálisis o la síntesis de moléculas, la preparación de compuestos y los recubrimientos duros se conocen desde hace tiempo y se basan en interacciones a nanoescala. Lo que sí es nuevo es que los sensores ultrasensibles nos dan la capacidad de comprender estos fenómenos mejor. Las nuevas técnicas de síntesis —de membranas, de nanocristales y de nuevos materiales— deberían contribuir a mejorar la tecnología allí donde la sociedad más lo necesita: en células de combustible, almacenamiento de energía y control de la contaminación. En el campo de las ciencias de la vida, el uso de nanoherramientas y el desarrollo de la nanotecnología están dando aún sus primeros pasos. Para que pueda ser de más utilidad las herramientas deben ser de más fácil manejo, una tarea que requiere un nuevo diseño de los sistemas pero que haría posible descifrar de manera rápida y barata la genética de un organismo complejo, diagnosticar y también administrar medicamentos a través de nanosistemas encapsulados, con la contribución que ello supondría para la medicina preventiva.

Antes de terminar, volveremos a la cuestión de cómo están afectando a la sociedad los avances en ciencia e ingeniería y, específicamente, en nanotecnología. La conclusión a que llegamos es que los problemas tienen que ver sobre todo con el empeño de individuos e instituciones por «triunfar». En las ciencias de la vida ha habido una gran toma de conciencia por parte de la sociedad debida, entre otras razones, al protagonismo de las compañías farmacéuticas y a la relativa facilidad con que se

6

Véase F. Dysson, «The Future Needs Us!», en *New York Review of Books*, vol. 50, n.º 2, del 13 de febrero de 2003.

El comportamiento emergente, es decir, el comportamiento impredecible, aparece en sistemas complejos, a saber, en aquellos con un gran número de elementos en interacción. La conducta de masas es un ejemplo de ello y constituye un tema apropiado de debate y reflexión acerca de creaciones del hombre susceptibles de cobrar vida propia. Sin embargo, un ejemplo de esto inspirado en la nanotecnología y que atrajo considerable atención pública, el libro de Michael Crichton, *Presa*, está basado en un supuesto científico erróneo. Es especialmente poderosa la descripción que hace el novelista de ejércitos de nanorrobots que colonizan a seres humanos y a su entorno, que son capaces de volar y nadar a gran velocidad a la manera de insectos u otros organismos similares de mayor tamaño. Esto no es posible porque la fuerza de resistencia viscosa sobre el área de superficie aumentada ralentiza los objetos a nanoescala. Es como si un ser humano tratara de nadar en melaza.

7

R. McGinn, «Ethics and Nanotechnology: Views of Nanotechnology Researchers», *Nanoethics*, vol. II, n.º. 2, 2008.

8

Ver <http://pubs.acs.org/cen/news/86/i15/8615news1/html>. El impacto de los residuos de nitrógeno y fósforo procedentes del uso a gran escala de fertilizantes puede apreciarse en la mayor parte del mundo occidental. En el oriental, los problemas de las zonas más desfavorecidas son la desaparición de la capa freática (y su sustitución con agua de mar en las zonas costeras) y la perforación de pozos cada vez más profundos que alcanzan la capa contaminada de arsénico, como ocurre en Bengala occidental, en India, y en Bangladesh. Este agotamiento masivo de los recursos de agua fue consecuencia de la generalización del uso del motor diesel y eléctrico, un invento de dos siglos de antigüedad pero que llegó al Tercer Mundo recientemente.

producen calamidades (pensemos en el ántrax, el Vioxx, el tabaco o la taliamida). Los científicos que trabajan en los campos de la física y la ingeniería necesitan desarrollar nuevos enfoques de manera que los procesos de investigación y el desarrollo continúen siendo éticos y tengan el bienestar de la sociedad como fin último.

La investigación a nanoescala, debido a su gran potencial, sobre todo en los ámbitos de la salud y el medioambiente, requiere ser conducida de acuerdo a determinados principios éticos y puesta en práctica de manera responsable. Es aquí donde surgen algunos conflictos potenciales.

El bajo coste de conectarse para intercambiar información, la gran cantidad de información almacenada y la inclinación natural de los individuos y de los Estados a fisgonear constituyen una pesadilla potencial que se ha intensificado en los últimos años tanto en Occidente como en Oriente. La nanotecnología refuerza esta capacidad potencial. ¿Cómo deben la sociedad y la investigación resolver este problema?

La humanidad es la primera creación resultado de la evolución capaz de transformar la ley de la supervivencia del más fuerte. ¿Cuál es la relación entre el hombre y la naturaleza? ¿Deberíamos reconocer y aceptar la supremacía de la naturaleza? ¿Es éticamente aceptable o responsable alterar, reconstruir incluso, organismos naturales? Cuando reemplazamos o aumentamos partes de organismos vivos, ¿dónde está el límite entre el ser humano y la máquina? No pasará mucho tiempo antes de que sensores neurales nos desvelen los secretos de las emociones humanas, de la personalidad e incluso tal vez de la conciencia. En los tribunales de la India ya se empieza a aceptar la resonancia magnética como indicio de honestidad o fraude. En el laboratorio, se han aplicado actuadores de impulsos neuronales a monos de laboratorio, estimulándoles electrónicamente a que pelen un plátano. Los debates Joy-Dysson⁶ se centraron precisamente en los miedos que suscita la potencial llegada de esta nueva tecnología antes de que la sociedad esté preparada para manejarla de manera segura y ecuánime.

¿Debería interrumpirse entonces la investigación relacionada con la nanotecnología debido a estas cuestiones controvertidas o a los desastres potenciales que podría acarrear, tal como han sugerido algunos? Nuestra opinión es que todos los puntos conflictivos deben ser identificados y cuidadosamente considerados. Pero ello no debe impedir que prosigan la investigación y el desarrollo, que deberán, no obstante, tener siempre la práctica responsable como lema principal. Igualmente importante es proporcionar oportunidades para los que trabajarán en este campo en el futuro y que hoy todavía son estudiantes, de manera que puedan reflexionar sobre su labor en el contexto de la sociedad en que se desarrolla y perfila.

La importancia de la seguridad en la manipulación y el uso de nanomateriales, dados los riesgos potenciales debidos a las propiedades reactivas de la nanoescala, se

ha puesto de manifiesto en un estudio reciente conducido entre investigadores de este campo.⁷ Las prácticas seguras están relacionadas con la cultura dentro del laboratorio y en una comunidad en general. Consideraciones de tiempo, dinero, estatus y competencia pueden inducir a investigadores y gerentes a rebajar sus exigencias. Históricamente en casi todas las áreas los gobiernos han hecho lo mínimo en lo referente a seguridad hasta que las circunstancias les han obligado. La regulación de seguridad ha sido la asignatura pendiente en la minería, las fábricas textiles, de neumáticos, asbestos, glicol y otros agentes químicos de la industria de los semiconductores, así como el plomo en la pintura y la gasolina. Aún seguimos debatiendo el posible papel de los teléfonos móviles en el aumento de la incidencia de tumores cerebrales debido a las interacciones electromagnéticas, casi una década después de que su uso se haya generalizado. Muchos que desempeñan puestos de responsabilidad siguen sin reconocer la responsabilidad de los seres humanos y de las emisiones de gases de efecto invernadero en el calentamiento global. Aunque es muy probable que la nanotecnología llegue a ser una importante herramienta para prevenir la contaminación, por ejemplo, facilitando la retirada de arsénico y la limpieza de las aguas, los nanomateriales también son potencialmente contaminantes. La plata se emplea como agente antibacteriano en la industria farmacéutica (en tiritas, etcétera) y también en la lucha contra los microbios. ¿Qué proporción del material empleado está penetrando el flujo de agua?⁸ Dadas las grandes cantidades de dinero que se están invirtiendo en desarrollo, las presiones para que se aprueben medidas reguladoras del medioambiente sin la investigación científica previa necesaria serán intensas. Aunque estos riesgos son el resultado de la ausencia de procedimientos y procesos sociales adecuados y no de la nanotecnología per se, los investigadores de este campo deberán tenerlos en cuenta.

Porque tratar de prevenir el daño mientras llevan a cabo su empresa científica es una responsabilidad ética fundamental de los científicos e ingenieros. Más allá de fomentar la seguridad en el laboratorio, preservar la integridad de los datos manejados, reconocer debidamente las contribuciones y respetar los derechos de propiedad intelectual, ¿tiene el investigador responsabilidades sobre las consecuencias sociales de su investigación? Las bombas atómicas en Hiroshima y Nagasaki durante la Segunda Guerra Mundial inauguraron un largo periodo de introspección y de compromiso público de los científicos con el debate social,⁹ porque ¿cómo se controla al genio una vez ha salido de la botella? De acuerdo a la visión más tradicional, la sociedad en general y no el investigador individual, es moralmente responsable de lo que se hace con conocimiento público. Sin embargo, el investigador individual con conocimiento de causa también tiene parte de responsabilidad. Los investigadores no siempre pueden declararse ignorantes de los riesgos que suponen las potentes

9

La ciencia y los científicos rara vez sirven de inspiración al arte. *Doctor Atómico*, una ópera popular estrenada en 2005, estaba inspirada en el proyecto Manhattan, donde se hizo realidad la bomba atómica. Los principales científicos del país entonces, capitaneados por Robert Oppenheimer, debatían acerca de la bomba mientras trabajaban sin pausa en lo que Oppenheimer, personaje central de la obra, llamó después del ensayo y citando a Bhagwat Heeta: «Soy la Muerte, el destructor de mundos». Es interesante señalar que en los bombardeos de la Segunda Guerra Mundial murieron muchas más personas inocentes que con la bomba atómica. Más tarde, en guerras abiertas y encubiertas, el Agente Naranja, el uranio empobrecido, las bombas de racimo y las minas antipersona han sembrado igualmente la muerte sin desencadenar una reacción similar en la opinión pública, probablemente a causa de que sus efectos perniciosos son más espaciados en el tiempo.

«máquinas» que crean. Los investigadores contemporáneos desarrollan y facilitan la difusión de sus creaciones dentro de sociedades que conocen. Aunque no siempre se puede predecir que determinados frutos de la investigación resulten en aplicaciones éticamente dudosas, en algunos casos sí es posible, por ejemplo, cuando están en juego sustanciosos beneficios potenciales militares y económicos. Por tanto, si un investigador tiene razones para sospechar que su trabajo pueda ser puesto en práctica en la sociedad con riesgo para los seres humanos tiene el deber moral de alertar a las autoridades pertinentes o al público.

Los ejemplos expuestos y la breve discusión sobre la cuestión de las responsabilidades ponen de manifiesto las dificultades surgidas cuando los avances científicos con implicaciones sociales de magnitud se producen con rapidez, y la sociedad se ve obligada a encontrar el equilibrio entre fomentar la investigación y el desarrollo productivos y defender un marco regulatorio que garantice la seguridad. Una posible respuesta a este desafío es que, en los

últimos años, la presión de la opinión pública ha servido para que el científico desarrolle una doble faceta: la de virtuoso técnico y detector de las implicaciones éticas de su trabajo. En palabras de Samuel Johnson, «la integridad sin conocimiento es débil y carece de utilidad, y el conocimiento sin integridad es peligroso y terrible».

Para los científicos e ingenieros, uno de los placeres de su disciplina es que la gran ciencia siempre es fuente de júbilo, de satisfacción emocional por el hecho de haber descubierto una porción de verdad que pueden considerar como propia. La ingeniería creativa proporciona el placer de combinar el descubrimiento científico con la alegría de haber hecho una aportación positiva a la humanidad. En su máxima representación, estas empresas encarnan los ideales de la vida civilizada, la búsqueda de la verdad y la práctica de la ciudadanía responsable. La nanotecnología forma parte de esta tradición clásica; está aquí, crece poderosamente y, convenientemente administrada, impulsará a la sociedad hacia adelante de muchas y provechosas maneras.

historia de la informática

PAUL E. CERUZZI

No hay palabra de la que se haya abusado más al hablar de informática que «revolución». Si creemos lo que dicen la prensa diaria y la televisión, cada modelo nuevo de chip, cada componente nuevo de *software*, cada nuevo adelanto en las redes sociales y cada modelo nuevo de teléfono móvil u otro dispositivo portátil cambiarán nuestra vida de forma revolucionaria. Unas semanas más tarde el objeto de esos reportajes curiosamente queda olvidado y pasa a sustituirse por un nuevo avance, el cual, se nos asegura, constituye, esta vez sí, el verdadero punto de inflexión.

Sin embargo es indiscutible que el efecto de la tecnología informática en la vida diaria del ciudadano de a pie ha sido revolucionario. Sólo con medir la capacidad de cálculo de estas máquinas, tomando como referencia la cantidad de datos que pueden almacenar y recuperar de su memoria interna, se pone de manifiesto un ritmo de progreso que ninguna otra tecnología, ni antigua ni moderna, ha alcanzado. No hace falta recurrir a los lenguajes especializados de ingenieros o programadores informáticos, pues la enorme cantidad de ordenadores y aparatos digitales que hay instalados en nuestros hogares y oficinas o que los consumidores llevan de un lado a otro por todo el mundo revela un ritmo de crecimiento parecido y que no da muestras de estar aminorando. Una medida aún más significativa nos la proporciona lo que estas máquinas son capaces de hacer. El transporte aéreo comercial, la recaudación de impuestos, la administración e investigación médica, la planificación y las operaciones

militares; estas y muchísimas otras actividades llevan el sello indeleble del apoyo informático, sin el cual serían muy diferentes o, sencillamente, no existirían.

Al intentar resumir la historia de la informática a lo largo de las últimas décadas nos enfrentamos a la dificultad de escribir en medio de esta fulgurante evolución. Si queremos hacerlo con el rigor debido, habremos de reconocer que tiene sus raíces históricas en la base de la civilización, que en parte se ha caracterizado por la capacidad de las personas de manejar y almacenar información por medio de símbolos. Pero en ella también debemos recoger los rápidos avances y la difusión vertiginosa de que ha sido objeto desde 1945, lo que no es fácil, si queremos conservar simultáneamente la perspectiva histórica. Este artículo es un breve repaso de las personas, las máquinas, las instituciones y los conceptos fundamentales que constituyen la revolución informática tal y como la conocemos en la actualidad. Empieza con el ábaco —que además del primero por orden alfabético es, cronológicamente, uno de los primeros instrumentos de cálculo— y llega hasta el siglo **xxi**, en el que las redes de ordenadores personales se han convertido en algo habitual y en el que la potencia informática ha terminado por integrarse en minúsculos dispositivos portátiles.

Aunque los aparatos digitales siguen evolucionando a mayor velocidad que nunca, los ordenadores personales se han estancado. Sus componentes físicos se han estabilizado: un teclado (procedente de la famosa máquina de

escribir de la década de 1890); una caja rectangular que contiene los circuitos electrónicos y la unidad de almacenamiento, y encima de ella, un terminal de visualización (heredero de la ya mítica pantalla de televisión de finales de la década de 1940). Lo mismo ha ocurrido con los circuitos electrónicos que hay en su interior, al margen de que cada año tengan mayor capacidad: durante los últimos treinta y cinco años han estado compuestos de circuitos integrados de silicio revestidos de tubos de plástico negro montados en paneles también de plástico. Los ordenadores portátiles dieron al traste con esta configuración, pero esencialmente son iguales. Tanto ingenieros como usuarios están de acuerdo en que su diseño físico presenta numerosos inconvenientes. Pensemos, por ejemplo, en las lesiones de los músculos de las manos que se producen por el uso excesivo de un teclado que se diseñó hace un siglo. Ahora bien, todavía no ha tenido éxito ninguno de los muchos intentos por lograr una potencia, una versatilidad y una facilidad de uso equivalentes en otras plataformas, en especial en teléfonos portátiles.

Los programas que estos ordenadores ejecutan, el *software*, continúan evolucionando a gran velocidad, como también lo hacen los elementos a los que están conectados, las bases de datos y las redes mundiales de comunicaciones. Es imposible prever adónde nos llevará todo ello. En el lapso de tiempo que transcurrirá desde la redacción de este ensayo hasta su publicación, puede que la naturaleza de la informática haya cambiado tanto que algunas partes de este estudio habrán quedado obsoletas. Los ingenieros de Silicon Valley hablan de que los avances en informática se desarrollan en *tiempo Internet*, unos seis años más rápido de lo que lo hacen en cualquier otro lugar. Incluso tras eliminar parte de esta hipérbola publicitaria, esta observación parece ser cierta.

Los orígenes de la informática pueden situarse al menos en cuatro momentos históricos. El primero es el más obvio: la Antigüedad, cuando civilizaciones nacientes empezaron a ayudarse de objetos para calcular y contar tales como las piedrecillas (en latín *calculi*, del que viene el término actual *calcular*), los tableros de cálculo y los ábacos, todos los cuales han llegado hasta el siglo xx (Aspray 1990).

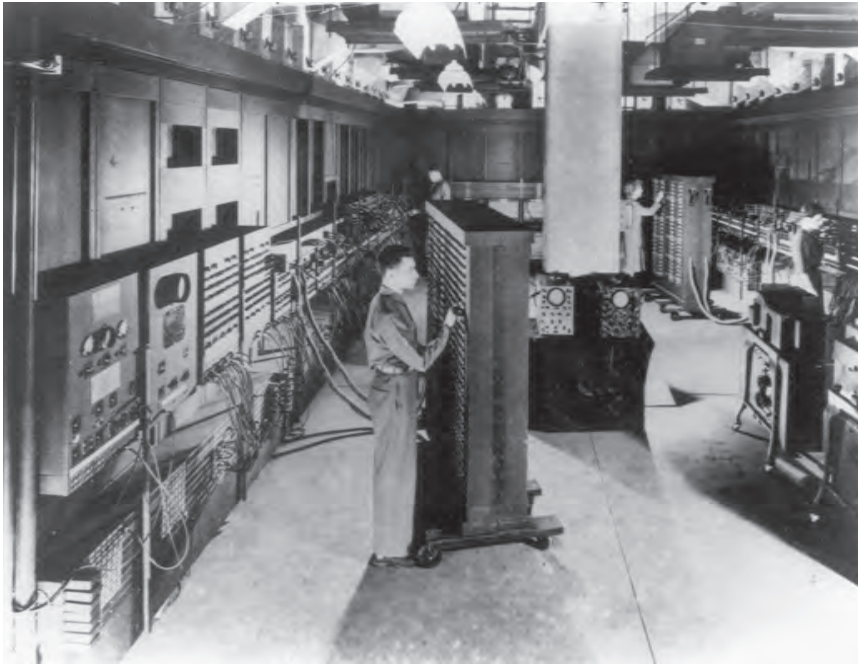
Ahora bien, ninguno de estos instrumentos se parece a lo que hoy nos referimos con el término *ordenador*. Para los ciudadanos de la época actual, un ordenador es un dispositivo o conjunto de dispositivos que nos libera de la pesadez que suponen las tareas de cálculo, así como de la actividad paralela de almacenar y recuperar información. Por tanto, el segundo hito histórico en la historia de la informática sería 1890, año en el que Herman Hollerith concibió la tarjeta perforada junto con un sistema de máquinas que procesaban, evaluaban y clasificaban la información codificada en ellas para la elaboración del censo de Estados Unidos. El sistema de Hollerith surgió en un momento crucial de la historia: cuando la maquinaria mecánica, cuyo mayor exponente son el motor de vapor y las turbinas

hidráulicas y de vapor, había transformado la industria. La conexión entre energía y producción hacía necesaria una mayor supervisión, no sólo física, también de la gestión de datos que la industrialización trajo consigo. Los tabuladores de Hollerith (y la empresa que éste fundó y que sería la base del grupo IBM) fueron una de tantas respuestas, pero hubo otras, como las máquinas eléctricas de contabilidad, las cajas registradoras, las máquinas de sumar mecánicas, la conmutación automática y los mecanismos de control para los ferrocarriles, las centrales telefónicas y telegráficas junto con los sistemas de información para los mercados internacionales de valores y materias primas.

No obstante, el lector actual podría quejarse y aducir que éste tampoco es el punto de partida adecuado. Parece que la auténtica revolución informática guarda relación con la electrónica, si no con los microprocesadores de silicio, que en la actualidad están en todas partes, al menos con sus antepasados inmediatos, los transistores y los tubos de vacío. Según esto, la era de la informática comenzó en febrero de 1946, cuando el ejército de Estados Unidos hizo público el Calculador e integrador numérico electrónico (Electronic Numerical Integrator and Computer, ENIAC) en un acto celebrado en la Moore School of Electrical Engineering de Filadelfia. El ENIAC, que contaba con 18.000 tubos de vacío, se presentó como un instrumento capaz de calcular la trayectoria de un proyectil lanzado desde un cañón antes de que el proyectil realizara el recorrido. Eligieron muy bien el ejemplo, pues este tipo de cálculos era el motivo por el cual el ejército había invertido más de medio millón de dólares de entonces (lo que equivaldría a varios millones de dólares en la actualidad) en una técnica que, se reconocía, era arriesgada y estaba por demostrar.

Un estudio histórico reciente ha desvelado que previamente existía otra máquina que realizaba operaciones de cálculo con tubos de vacío. Se trata del *Colossus* británico, del que se fabricaron varias unidades que se instalaron en Bletchley Park, Inglaterra, durante la Segunda Guerra Mundial, y se usaron con éxito para descifrar los códigos alemanes. A diferencia del ENIAC, estas máquinas no realizaban operaciones aritméticas convencionales, pero sí llevaban a cabo operaciones de lógica a gran velocidad, y al menos algunas de ellas llevaban varios años en funcionamiento antes de la presentación pública del invento estadounidense. Tanto el ENIAC como el *Colossus* estuvieron precedidos de un dispositivo experimental que diseñó en la Universidad de Iowa un catedrático de Física llamado John V. Atanasoff, con la colaboración de Clifford Berry. Esta máquina también realizaba operaciones de cálculo por medio de tubos de vacío, pero, aunque sus componentes principales se presentaron en 1942, nunca llegó a estar en funcionamiento (Burks y Burks 1988).

El lector podría observar de nuevo que lo fundamental no es simplemente que una tecnología exista, sino que pase a ser de uso habitual en las mesas de trabajo y los



El ENIAC, Universidad de Pensilvania, 1945. Smithsonian Institution.

hogares del ciudadano normal. Después de todo no han sido muchas las personas, como máximo una docena, que hayan tenido la oportunidad de utilizar el ENIAC y sacar provecho de su extraordinaria potencia. Lo mismo ocurre con los ordenadores *Colossus*, que se desmontaron después de la Segunda Guerra Mundial. Según esto, habría que fechar el *verdadero* origen de la revolución informática no en 1946 sino en 1977, año en el que dos jóvenes, Steve Jobs y Steve Wozniak, originarios de lo que se conoce como Silicon Valley, dieron a conocer al mundo un ordenador llamado Apple II. El Apple II (al igual que su predecesor inmediato el Altair y su sucesor el IBM PC) sacó a la informática del mundo especializado de las grandes empresas y el ejército y la llevó al resto del mundo.

Podríamos seguir indefinidamente con este debate. Según los jóvenes de hoy, la revolución informática es aún más reciente, pues consideran que se produjo cuando, gracias a Internet, un ordenador en un lugar determinado intercambiaba información con ordenadores que estaban en otros lugares. La más famosa de estas redes la creó la Agencia de proyectos de investigación avanzada (Advanced Research Projects Agency, ARPA) del Departamento de Defensa de Estados Unidos, que a principios de 1969 ya tenía una red en marcha (ARPANET). Sin embargo, también hubo otras redes que conectaron ordenadores personales y miniordenadores. Cuando éstas se combinaron, en la década de 1980, nació Internet tal y como hoy la conocemos (Abbate 1999).

Lo cierto es que hay muchos puntos donde se puede empezar esta historia. Mientras escribo este artículo la informática está experimentando una nueva transformación. Me refiero a la fusión entre ordenadores perso-

nales y dispositivos de comunicación portátiles. Como en otras ocasiones, esta transformación viene acompañada de descripciones en la prensa diaria que hablan de los efectos *revolucionarios* que tendrá. Es evidente que el teléfono posee una historia larga e interesante, pero no es éste el tema que nos ocupa. Sólo hay una cosa clara: aún no hemos asistido al último episodio de este fenómeno. Habrá muchos más cambios en el futuro, todos impredecibles, todos presentados como el último adelanto de la revolución informática y todos dejarán relegadas al olvido las «revoluciones» anteriores.

Este relato comienza a principios de la década de 1940. La transición de los ordenadores mecánicos a los electrónicos fue, en efecto, importante, pues entonces se sentaron las bases para inventos posteriores, como los ordenadores personales. En aquellos años ocurrieron más cosas importantes: fue durante esta década cuando surgió el concepto de *programación* (posteriormente ampliado al de *software*) como actividad independiente del diseño de los equipos informáticos, si bien de suma importancia para que éstos pudieran emplearse para lo que habían sido diseñados. Por último, fue en esta época cuando, como resultado de la experiencia con las primeras enormes computadoras experimentales ya en funcionamiento, apareció un diseño funcional básico, una *arquitectura*, para utilizar el término más reciente, que se ha mantenido a través de las oleadas sucesivas de avances tecnológicos hasta la actualidad.

Por tanto, y con todos los matices que habrá que añadir para que la afirmación resulte admisible para los historiadores, podemos considerar que el ENIAC constituyó el eje de la revolución informática (Stern 1981). Aquella máquina, concebida y desarrollada en la Universidad de Pensilvania durante la Segunda Guerra Mundial, inauguró lo que conocemos por *era informática*. Siempre y cuando se entienda que cualquier punto de origen histórico que se elija es en cierto modo arbitrario, y siempre y cuando se conceda el debido crédito a los adelantos que tuvieron lugar antes, incluida la labor de Babbage y Hollerith, así como los inventos de la máquina de sumar, la caja registradora y otros dispositivos similares, podemos empezar aquí.

Introducción

Casi todas las culturas han compartido la capacidad de contar y de representar cantidades con notaciones simbólicas de algún tipo, por muy *primitivas* que puedan parecerles a los estudiosos actuales. Ahora bien, conseguir pruebas materiales de ello es mucho más difícil, a menos que utilizarasen materiales duraderos como las tablillas de arcilla. Sabemos que la idea de representar y manejar información cuantitativa de manera simbólica con piedrecillas, cuentas, nudos en una cuerda o métodos similares surgió de manera independiente en todo el mundo antiguo. Por ejemplo, los exploradores españoles en el Nuevo Mundo descubrieron que los incas utilizaban un avanza-

do sistema de cuerdas con nudos llamado *quipu*, y que en la Biblia se menciona un sistema parecido de sartas con nudos, y que al menos una de ellas, el rosario, ha sobrevivido hasta nuestros días. Un modelo de representación muy abstracto de las cuentas evolucionó en el ábaco, del que como mínimo han llegado hasta la actualidad tres tipos diferentes en China, Japón y Rusia, y que en manos de un operador diestro constituye una herramienta de cálculo potente, compacta y versátil. En la Edad Media los países occidentales también utilizaron asistentes de cálculo parecidos, entre ellos unos tableros (dotados de cuadrículas y patrones para facilitar las sumas) y sus fichas (que han llegado hasta nosotros en forma de las fichas de juego empleadas en los casinos).

Es importante señalar que estos instrumentos sólo los usaban aquellas personas cuyos cargos dentro del gobierno, la Iglesia o los negocios lo requiriesen. Hecha esta salvedad podría decirse que eran de uso *común*, aunque no en el sentido de que estuvieran en todas partes. Esta misma salvedad se puede aplicar a todas las máquinas de cálculo, ya que su adopción depende, sin duda, de lo costosas que sean, si bien resulta además fundamental que se ajusten a las necesidades de quienes las van a usar. Cuando la sociedad occidental se industrializó y se volvió más compleja, estas necesidades aumentaron; no obstante conviene apuntar que a pesar de lo mucho que han bajado los precios de los ordenadores y del acceso a Internet, no se ha conseguido aún que penetren completamente en el mercado del consumidor y, probablemente, nunca lo hagan.

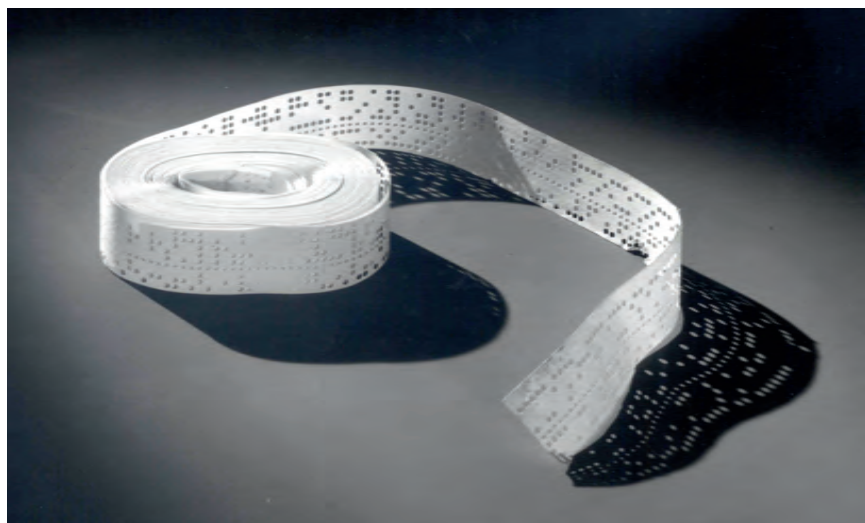
Antes de pasar a las máquinas conviene mencionar otra herramienta de cálculo que tuvo un uso muy extendido y que ha llegado hasta la época moderna de forma muy rudimentaria. Se trata de las tablas impresas, en las que había, por ejemplo, una lista de valores de una función matemática determinada. Su uso data de la Grecia antigua, pero

también las utilizaron mucho los astrónomos y, aún más importante, los marinos en alta mar. El negocio de los seguros, por su parte, desarrolló las llamadas tablas de estadísticas, como las de los índices de mortalidad, por ejemplo. En la actualidad, las calculadoras de bolsillo y las hojas de cálculo de los programas informáticos nos permiten realizar operaciones de manera inmediata, pero las tablas todavía tienen su valor. Aún es posible encontrar lugares donde se utilizan, lo que pone de manifiesto su estrecha relación con uno de los usos principales de los modernos instrumentos de cálculo electrónico (Kidwell y Ceruzzi 1994).

La mayoría de estos instrumentos funcionaban en colaboración con el sistema de numeración indo-árabe, en el que el valor de un símbolo depende no sólo del símbolo en sí (1, 2, 3...), sino también del lugar donde está situado (y en el que el importantísimo cero se usa como un parámetro de sustitución). Este sistema de numeración era mucho más avanzado que los de tipo aditivo, como el romano, y su adopción por parte de los europeos a finales de la Edad Media constituyó un hito en el camino hacia el cálculo moderno. Cuando realizaban operaciones de suma, si el total de los dígitos de una columna era superior a nueve había que *llevarlo* a la siguiente columna por la izquierda. La mecanización de este proceso supuso un paso significativo desde las ayudas de cálculo mencionadas anteriormente hacia el desarrollo del cálculo automático. Una descripción esquemática y fragmentaria recogida en una carta a Johannes Kepler revela que el profesor Wilhelm Schickard, de la localidad alemana de Tübinga, había diseñado un aparato de estas características a principios del siglo XVII, pero no hay constancia de que ninguna de las piezas haya llegado hasta nuestros días.

En 1642 el filósofo y matemático francés Blaise Pascal inventó una máquina de sumar que es la más antigua de cuantas se conservan. Los dígitos se introducían en la calculadora haciendo girar un conjunto de ruedas, una por cada columna. Cuando las ruedas superaban el 9, un diente del engranaje avanzaba una unidad en la rueda contigua. Pascal se esforzó mucho para asegurarse de que el caso extremo de sumar un 1 a una serie de 9 no bloquease el mecanismo. Esta máquina inspiró a unos cuantos inventores a construir aparatos parecidos, pero ninguno se comercializó con éxito. Ello se debió, por un lado, a que eran frágiles y delicados y, por lo tanto, costosos y, por otro, a que en la época de Pascal no se consideraba que estas máquinas fueran necesarias.

Unos treinta años más tarde el filósofo y matemático alemán Gottfried Wilhelm Leibniz, satirizado por Voltaire en su *Cándido* y famoso por ser uno de los creadores del Cálculo, tuvo noticias del invento de Pascal e intentó diseñar una calculadora. Consiguió construir una máquina que no sólo sumaba sino también multiplicaba mediante el uso de engranajes que conectaban un número variable de dientes dependiendo de dónde hubiera puesto el operador el círculo indicador. Esta calculadora no funcionó bien, pero el *tambor*



Cinta de papel BASIC, 1975. Casi tan pronto como tuvo noticias del Altair, Bill Gates abandonó sus estudios y se trasladó a Albuquerque con Paul Allen, un amigo del instituto. Ambos crearon la versión del lenguaje de programación BASIC para el Altair, que fue crucial para que este ordenador se convirtiese en un dispositivo viable. Smithsonian Institution.

escalonado se convirtió en la base para casi todas las calculadoras de multiplicar hasta finales del siglo XIX. Uno de sus descendientes modernos, el Curta, era lo suficientemente pequeño como para que cupiese en un bolsillo, y se fabricó y comercializó hasta comienzos de la década de 1970.

La aparición de una sociedad más mercantil, con una clase media en aumento, contribuyó a hacer más favorables las condiciones para el éxito comercial. Hacia 1820, Charles Xavier Thomas, precursor de la estabilización del sector de seguros en Francia, diseñó y comercializó su *Aritmómetro*, en el que utilizó el tambor escalonado de Leibniz para hacer multiplicaciones. Al principio no se vendieron muchos, pero después de 1870 su uso se extendió y llegaron a venderse unos cien ejemplares al año. Para entonces la industrialización estaba en pleno desarrollo y, junto a la máquina de Thomas, hicieron su aparición una serie de productos que compitieron entre sí para satisfacer la creciente demanda (Eames y Eames 1990).

Esto ocurrió a ambos lados del Atlántico. Son especialmente importantes dos *máquinas de sumar* desarrolladas en Estados Unidos. Aunque ninguna de las dos podía multiplicar, sí efectuaban sumas a gran velocidad, eran fáciles de usar, su coste era modesto (aunque no bajo) y eran muy resistentes, lo que las hacía rentables. A mediados de la década de 1880, Dorr E. Felt diseñó y patentó una máquina de sumar que se usaba presionando un conjunto de teclas numéricas, una serie de dígitos por cada posición numérica. Y, lo que era más novedoso aún, la fuerza necesaria para presionar las teclas activaba el mecanismo de manera que el operador no tenía que detenerse y girar una manivela, tirar de una palanca o ninguna otra cosa. En manos de un operador diestro, que no separase los dedos del teclado, ni siquiera lo mirase, el *Comptómetro* de Felt podía realizar sumas con enorme rapidez y precisión. Con un precio de venta de unos 125 dólares, los *comptómetros* pronto se convirtieron en una herramienta habitual en las oficinas estadounidenses de principios del siglo XX. Por la misma época, William Seward Burroughs inventó una máquina de sumar que imprimía los resultados en una tira de papel y evitaba tener que consultarlos en la ventanilla. Su invento supuso el comienzo de Burroughs Adding Machine Company, que en la década de 1950 hizo el tránsito a la fabricación de ordenadores electrónicos, y que tras una fusión con Sperry en 1980 se conoce con el nombre de Unisys Corporation.

En las oficinas de Europa las máquinas de calcular también se convirtieron en un producto de uso habitual, aunque tomaron un camino diferente. El ingeniero sueco W. Odhner inventó una máquina compacta y sólida que multiplicaba además de sumar, mediante un tipo de engranaje diferente al de Leibnitz (los números se introducían activando palancas en lugar de presionando teclas), y que se comercializó con éxito con los nombres de Odhner, Brunsviga y otros.

No se puede dar por concluido ningún estudio sobre máquinas de cálculo sin mencionar a Charles Babbage,

un británico a quien muchos consideran el inventor del primer ordenador automático y programable, la famosa *máquina analítica*. Esta idea se le ocurrió tras diseñar y montar parcialmente una *máquina diferencial*, un proyecto más modesto pero que representaba un gran avance para la tecnología de cálculo de su época. Más adelante hablaremos en detalle de la labor de Babbage; baste decir ahora que lo que presentó, a principios de la década de 1830, era el proyecto de una máquina con todos los componentes básicos funcionales de un ordenador moderno: una unidad aritmética que llamó *Mill*, un dispositivo de memoria que llamó *Store*, un método de programar la máquina por medio de tarjetas y una forma de imprimir los resultados o perforar las respuestas en otra serie de tarjetas. Se fabricaría con metal y funcionaría con un motor de vapor. Babbage pasó muchos años intentando llevar su idea a buen puerto, pero cuando murió en 1871 sólo se habían construido algunas partes.

Es curioso pensar en lo diferente que el mundo habría sido si Babbage hubiera logrado terminar su máquina. Quizás habríamos conocido una era de la información con motores de vapor. Sin embargo, como ya ocurriera con las máquinas de Pascal y Leibniz, hay que tener en cuenta que el mundo entonces no estaba necesariamente preparado para este tipo de invento. Para que hubiera tenido verdadera repercusión, Babbage no sólo habría tenido que superar los obstáculos técnicos que malograron su motor analítico, también desplegar unas dotes comerciales considerables para convencer a la gente de que su invento era realmente útil. La prueba de ello está en el hecho de que los suecos Georg Scheutz y su hijo Edvard finalizaron el diseño de una máquina diferencial operativa en 1853, considerada la primera calculadora con impresora de uso comercial (Merzbach 1977). Aunque el observatorio de Dudley de Albany, en el estado de Nueva York, la adquirió, lo cierto es que la máquina apenas tuvo repercusiones en la ciencia o el comercio. La era de la información aún tenía que esperar.

Hacia finales del siglo XIX el arte de calcular se había estabilizado. En el mundo de los negocios el sencillo *comptómetro* o el *Odhner* habían ocupado su lugar junto a otros aparatos de alcance similar, como la máquina de escribir o el teletipo. En el mundo de la ciencia, todavía pequeño en aquellos años, había cierto interés, pero no el suficiente para apoyar la fabricación de algo que fuera más allá de una máquina especializada aquí y allá. Las ciencias que necesitaban realizar cálculos, como la astronomía, se las arreglaban con las tablas impresas y las *calculadoras* humanas (así se llamaban quienes realizaban esta tarea) que trabajaban con papel y lápiz, libros de tablas matemáticas y, quizás, alguna máquina de sumar. Lo mismo ocurría con los ingenieros, utilizaban libros de tablas matemáticas ayudados en algunos casos por máquinas especializadas diseñadas para resolver un problema concreto (por ejemplo, un instrumento para pronosticar mareas o el analizador diferencial de Bush). A partir de 1900 los ingenieros también contaron

con la ayuda de dispositivos analógicos como el planímetro y, sobre todo, la regla deslizante, un instrumento de una precisión limitada, pero versátil, que satisfacía razonablemente la mayoría de las necesidades de los ingenieros.

Las tarjetas perforadas de Herman Hollerith empezaron como uno de estos sistemas especializados. En 1889 atendió a una petición del superintendente del censo de Estados Unidos, a quien cada vez le resultaba más difícil presentar sus informes en el momento debido. La tarjeta perforada, junto con el método de codificación de datos por medio de patrones de agujeros en esta tarjeta, y de clasificación y recuento de los totales y los subtotales que la acompañaban se ajustaba a la perfección a las necesidades de la Oficina del Censo. Lo que ocurrió después se debió sobre todo a la iniciativa de Hollerith quien, tras haber inventado este sistema, no se conformó con tener un único cliente que lo utilizase una vez cada diez años, por lo que inició una campaña para convencer a otros de su utilidad. Fundó una empresa, que en 1911 se fusionó con otras dos para constituir la Computing-Tabulating-Recording Corporation, y en 1924, cuando Thomas Watson tomó las riendas, pasó a llamarse International Business Machines. Watson, como vendedor que era, comprendió que estos aparatos tenían que satisfacer las necesidades de los clientes si querían prosperar. Entretanto, la Oficina del Censo, que no quería depender demasiado de un solo proveedor, fomentó el crecimiento de una empresa de la competencia, Remington Rand, que se convirtió en el rival principal de IBM en este tipo de equipos durante los cincuenta años que siguieron.

Visto en retrospectiva, da la impresión de que el éxito de los sistemas de tarjetas perforadas vino dictado de antemano, pues su capacidad para clasificar, recopilar y tabular información encajó a la perfección con la creciente demanda de datos relativos a las ventas, el marketing y la fabricación procedentes de una economía industrial en auge. No hay duda de que el factor suerte contribuyó, pero hay que conceder a Hollerith el crédito debido por su visión de futuro, al igual que a Watson por promocionar de manera incansable esta tecnología. Cuando en 1930 la economía de Estados Unidos se tambaleó, las máquinas IBM continuaron usándose tanto como antes, pues satisfacían el ansia de datos estadísticos de las agencias gubernamentales estadounidenses y extranjeras. Watson, vendedor por antonomasia, promovió y financió generosamente además posibles aplicaciones de los productos a su empresa en los ámbitos de la educación y la ciencia. A cambio de ello, algunos científicos descubrieron que los equipos IBM, con unas modificaciones mínimas, servían para resolver problemas científicos. Para astrónomos como L. J. Comrie la tarjeta perforada se convirtió, en efecto, en el sueño fallido de Babbage llevado a la práctica. Otros científicos, entre ellos el ya mencionado Atanasoff, habían empezado a diseñar calculadoras especializadas capaces de realizar una secuencia de operaciones, como se suponía habría hecho la máquina

analítica que Babbage nunca llegó a completar. Todos ellos lo consiguieron con la ayuda de los tabuladores y calculadoras mecánicas de IBM que cumplieron su función de forma tan satisfactoria que casi hicieron innecesario desarrollar un nuevo tipo de máquina (Eckert 1940).

Al revisar esta época se observa una correspondencia notable entre estos nuevos diseños de calculadoras programables y el de la máquina analítica que nunca llegó a completarse. Sin embargo, el único diseñador que conocía la existencia de Charles Babbage era Howard Aiken, un catedrático de la Universidad de Harvard, y ni siquiera él adoptó su modelo cuando desarrolló su propio ordenador. En 1930 Babbage no era un completo desconocido, pero la mayoría de las historias que sobre él circulaban coincidían en que su labor había sido un fracaso y sus máquinas, ideas descabelladas, lo cual no sirvió de gran inspiración a una nueva generación de jóvenes inventores. Sin embargo, todos los que tuvieron éxito donde Babbage fracasó compartían su pasión y determinación por llevar a la práctica, por medio de engranajes y cables, el concepto de cálculo automático. Además, también contaban con unas buenas dotes de persuasión, como las de Thomas Watson.

Entre ellos cabe mencionar a Konrad Zuse, quien mientras todavía cursaba sus estudios de Ingeniería en Berlín, a mediados de la década de 1930, hizo un esbozo de una máquina automática porque, según decía, era *demasiado perezoso* para efectuar las operaciones de cálculo necesarias para sus estudios. La pereza y la necesidad, dicen, son la madre de la ciencia. Cuando los nazis sumieron al mundo en la guerra, Zuse trabajaba durante el día en una planta aeronáutica en Berlín y por la noche construía máquinas experimentales en la casa de sus padres. En diciembre de 1941 puso en funcionamiento su Z3, utilizando relés telefónicos sobrantes para los cálculos y el almacenamiento, y películas fotográficas perforadas de desecho para la programación (Ceruzzi 1983).

En 1937 Howard Aiken se planteó, mientras trabajaba en su tesis de Física en Harvard, diseñar lo que más tarde se conoció como *Calculador controlado por secuencia automática* (Automatic Sequence Controlled Calculator, ASCC). Eligió las palabras deliberadamente con la intención de que reflejasen su opinión de que la falta de capacidad de las máquinas de tarjetas perforadas para efectuar secuencias de operaciones suponía una limitación para su uso científico. Aiken consiguió el apoyo de IBM, que construyó la máquina y la llevó a Harvard. Allí, en plena Segunda Guerra Mundial, en 1944, la dio a conocer. De ahí que el ASCC también se conozca por ser el primer invento que difundió la noción de cálculo automático (los espías alemanes comunicaron estas noticias a Zuse, pero para 1944 él ya tenía muy avanzada la construcción de una máquina de características similares a la de Aiken). El ASCC, o Harvard Mark I, como se le suele llamar, utilizaba componentes modificados IBM para los registros, pero se programaba por medio de una tira de papel perforado.

En 1937 George Stibitz, un matemático-investigador que trabajaba en los Bell Telephone Laboratories de Nueva York, diseñó un rudimentario circuito que efectuaba sumas por medio de la aritmética binaria, un sistema numérico difícil de usar para los seres humanos, pero que se adapta a la perfección a estos dispositivos. Al cabo de dos años consiguió convencer a su empresa para que fabricara una sofisticada calculadora a base de relés que funcionase con los llamados números *complejos*, que con frecuencia aparecían en los análisis de circuitos telefónicos. Esta calculadora de números complejos no era programable, pero contribuyó a la creación de otros modelos en los laboratorios Bell durante la Segunda Guerra Mundial que sí lo fueron. Todo ello culminó con el diseño de varios ordenadores de uso general de gran tamaño basados en relés, que tenían la capacidad no sólo de ejecutar una secuencia de operaciones aritméticas, sino también de modificar su forma de proceder basándose en los resultados de un cálculo previo. Esta última característica, junto con la velocidad electrónica (de la que trataremos después), se considera una diferencia esencial entre lo que hoy conocemos como *ordenadores* y sus predecesores de menor capacidad, las *calculadoras* (en 1943 Stibitz fue el primero que utilizó la palabra *digital* para describir máquinas que realizaban cálculos con números discretos).

Para completar este estudio de máquinas cabe mencionar al Analizador diferencial que diseñó el MIT (Massachusetts Institute of Technology, Instituto tecnológico de Massachusetts) bajo la dirección del catedrático Vannevar Bush a mediados de la década de 1930. Esta máquina no realizaba cálculos *digitalmente*, para usar la expresión actual, pero funcionaba con un principio parecido al de los contadores de vatios *analógicos* que se pueden encontrar en las casas. En otros aspectos, el analizador de Bush era parecido a otras máquinas de las que ya se hemos hablado anteriormente. Al igual que otros precursores, Bush buscaba resolver un problema específico: analizar las redes de los generadores de corriente alterna y las líneas de transmisión. El Analizador diferencial estaba formado por un complejo ensamblaje de unidades de cálculo que se podían reconfigurar para resolver una variedad de problemas. Debido a las necesidades de la Segunda Guerra Mundial se montaron varias unidades de esta máquina, pero se destinaron a resolver problemas más urgentes. Una de ellas, la que se instaló en la Moore School of Electrical Engineering de Filadelfia, sirvió de inspiración para el ENIAC.

Todas estas máquinas utilizaban engranajes mecánicos, ruedas, palancas o relés para sus elementos de cálculo. Los relés son dispositivos eléctricos, pero el interruptor activa la corriente de manera mecánica, con lo que la velocidad de la operación tiene, en esencia, las mismas características que las de los dispositivos completamente mecánicos. Ya en 1919 se sabía que era posible diseñar un circuito a base de tubos de vacío capaz de realizar la conmutación

con mayor rapidez, al producirse ésta dentro del tubo por medio de una corriente de electrones de masa insignificante. Los tubos eran propensos a quemarse, ya que para funcionar requerían una gran potencia que a su vez era fuente de excesivo calor. Los incentivos para construir una máquina de cálculo a base de tubos no eran demasiados, a menos que las ventajas, en lo que a la rapidez se refiere, superasen estos inconvenientes.

A mediados de la década de 1930 John V. Atanasoff, catedrático de Física de la Universidad de Iowa, observó las ventajas de emplear circuitos de tubos para resolver sistemas de ecuaciones lineales. Las ecuaciones lineales se pueden encontrar en casi todas las ramas de la Física y su solución requiere realizar un gran número de operaciones de aritmética ordinarias pero conservando los resultados intermedios. En 1939 Atanasoff, con una modesta beca universitaria, comenzó a diseñar circuitos y para 1942 tenía listo un prototipo que funcionaba, a excepción de fallos intermitentes ocurridos en la unidad de almacenamiento intermedia. Por entonces Atanasoff se trasladó a Washington para trabajar en otros proyectos durante la guerra y no llegó a terminar su ordenador. En aquella misma época en Alemania, un compañero de Zuse llamado Helmut Schreyer diseñó circuitos de tubos que presentó como sustitutos de los relés que empleaba Zuse. Aunque esta propuesta constituyó la base de su tesis doctoral, al margen de unos cuantos paneles experimentales, no avanzó mucho en ella.

La primera vez en la que se aplicaron con éxito los tubos de vacío a la informática fue en Inglaterra, donde un equipo de personas encargadas de descifrar códigos diseñó, en el más absoluto secreto, una máquina que les ayudara a interpretar los mensajes militares transmitidos por radio de los alemanes. Es un ejemplo que ilustra a la perfección la necesidad de la velocidad que proporcionaba la electrónica, pues no sólo había un considerable número de combinaciones de teclas a tener en cuenta sino que también el contenido de los mensajes interceptados perdía valor militar según pasaba el tiempo y a menudo quedaba obsoleto transcurridos unos días. El primero de los llamados *Colossus* se terminó en 1943 (más o menos en la época que se empezó el ENIAC), y para el final de la guerra había diez en funcionamiento. La información relativa a estas máquinas sigue estando clasificada, incluso después de 65 años, pero se ha desvelado que aunque no realizaban operaciones aritméticas como lo hacían las calculadoras, sí podían realizar y, así lo hicieron, operaciones de lógica con información expresada en símbolos, lo cual constituye la base de los circuitos electrónicos de procesamiento actuales.

El ENIAC, diseñado en la Universidad de Pensilvania y presentado al público en febrero de 1946, sigue más la tradición de las máquinas que acabamos de ver que la de los ordenadores electrónicos de uso general que le siguieron. Se concibió, propuso y diseñó para resolver un proble-

ma específico: calcular las tablas de balísticas del ejército. Su arquitectura es un reflejo de lo que se requería para resolver ese problema, y ningún otro ordenador la ha imitado. Sólo se construyó uno, y aunque el final de la guerra hizo que la elaboración de estas tablas no fuera tan urgente, las necesidades militares fueron siempre determinantes para la existencia del ENIAC (se desconectó en 1955). En la década de 1940 la informática estaba avanzando en varios frentes. Los ejemplos ya mencionados son los más destacados, pero, detrás de ellos hubo un gran número de proyectos que, aunque de menor envergadura, fueron también significativos.

La metáfora de progreso lineal (por ejemplo, el uso de términos como «hito») para relatar la historia de la informática no es adecuada. Los adelantos que se produjeron en este campo durante la década de 1940 se parecían más a un ejército avanzando por un terreno accidentado. El ENIAC, en virtud del aumento drástico de la velocidad con la que realizaba las operaciones aritméticas, hizo que la función de *cálculo* de estas máquinas se colocara muy por delante de otras funciones, como el almacenamiento de datos o la producción de resultados, con lo que hubo que darse prisa para situar a éstas al mismo nivel. De todas estas funciones, el mayor obstáculo lo constituyó la función mediante la que se daba instrucciones al procesador. John Mauchly señaló: «Sólo se pueden efectuar cálculos a gran velocidad si se dan instrucciones a gran velocidad». Por tanto, los diseñadores del ENIAC vieron claramente que era necesario crear una unidad electrónica de almacenamiento de instrucciones interna. Todas las máquinas disponen de *software*: un conjunto de procedimientos que hacen posible usarlas. Antes de la electrónica, la velocidad de la maquinaria guardaba relación con la de los seres

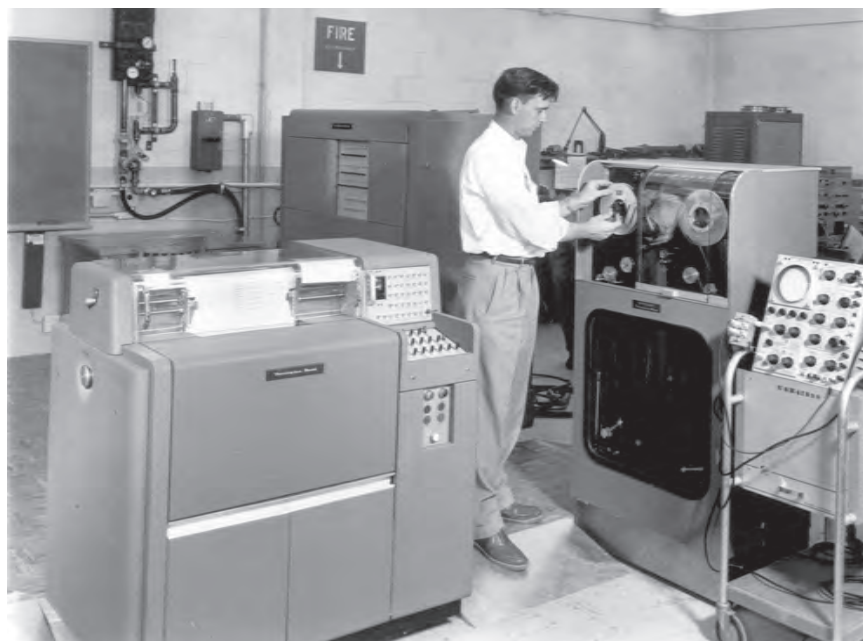
humanos. Esta separación aparece por primera vez con los ordenadores, y en ella reside la verdadera naturaleza *revolucionaria* de la era digital. El ENIAC, gracias a la elevada velocidad a la que efectuaba operaciones aritméticas, colocó la programación en primer plano (no es una coincidencia que la expresión «programar un ordenador» fuera acuñada por el equipo que diseñó el ENIAC).

El ENIAC, por tanto, ocupa un lugar paradójico, ya que constituye el eje de esta historia tanto por sus defectos como por sus virtudes. No estaba programado, sino que se *configuraba* de manera laboriosa conectando cables, que, en efecto, había que volver a conectar para cada nueva operación. Todo ello suponía un problema que se tardaba minutos en resolver, por lo que configurarlo podía llevar días. En cambio los parientes electromecánicos del ENIAC, como el Harvard Mark I, podían programarse en unas cuantas horas pero tardaban días en resolver las ecuaciones.

Cuando el ENIAC tomaba forma, a principios de la década de 1940, sus diseñadores estaban ya pensando en cómo sería la máquina que lo sucedería. En retrospectiva, se trataba de un equipo perfecto para la labor que tenía que realizar: personas con conocimientos de ingeniería eléctrica, de matemáticas y de lógica. De sus deliberaciones surgió la noción de diseñar un ordenador que contara con una unidad de memoria dedicada, que almacenase datos, pero que no necesariamente realizase operaciones aritméticas o de otro tipo. Las instrucciones, al igual que los datos, se almacenarían en este dispositivo, y cada uno de ellos se podría recuperar o almacenar a gran velocidad. Este requisito surgió de la necesidad práctica de ganar velocidad, como antes señaló Mauchly, así como del deseo de la ingeniería de disponer de una unidad de memoria simple sin la complicación adicional de tener que dividirla y asignar un espacio diferente para cada tipo de información.

De esta sencilla noción nació en gran medida la capacidad de cálculo que siguió y que desde entonces se ha asociado a John von Neumann, quien se unió al equipo del ENIAC y en 1945 escribió un informe sobre su sucesor, el EDVAC, explicando estos conceptos. Sin embargo, se trató claramente de un esfuerzo conjunto, que tuvo al ENIAC, entonces en proceso de montaje, como telón de fondo.

Todas las ventajas de este diseño no servirían de nada si no se encontraba un dispositivo de memoria con suficiente capacidad para operar de manera segura, rápida y barata. Eckert estaba a favor del uso de tubos de mercurio que transportaban impulsos acústicos, pero Newman prefería utilizar un tubo de vacío especial. Los primeros ordenadores que dispusieron de verdaderos programas en su memoria para su funcionamiento utilizaban tubos de mercurio o tubos de rayos catódicos modificados que almacenaban información a modo de haces de carga eléctrica (Randell 1975). Estos métodos proporcionaban alta velocidad, pero tenían una capacidad limitada y eran caros. Muchos otros ingenieros optaron por un tambor magnético rotativo que, aunque mucho más lento, era más segu-



El UNIVAC I, Laboratorio de Lawrence Livermore, California, 1952. Smithsonian Institution.



Macrocomputadora IBM System 360, 1965. El System 360 fue una de las macrocomputadoras más famosas y sentó las bases de la principal línea de fabricación de ordenadores hasta la década de 1990. Smithsonian Institution.

ro. El Proyecto Whirlwind, del MIT, superó este obstáculo cuando, a principios de la década de 1950, su equipo ideó una forma de almacenar datos en diminutos *núcleos* magnéticos, unas piezas de material imantado en forma de rosquilla (Redmond y Smith 1980).

Generaciones: 1950–1970

Eckert y Mauchly no sólo son famosos por sus contribuciones al diseño de ordenadores. Fueron de los pocos que, por aquella época, buscaron aplicaciones comerciales para su invento, en lugar de limitarse a usos científicos, militares o industriales a gran escala. Los británicos fueron los primeros en crear un ordenador para uso comercial: el LEO, una versión comercial del EDSAC diseñado para una empresa de *catering* llamada J. Lyons & Company Ltd., que estaba en funcionamiento en 1951. Pero al igual que ocurrió con los inventos de Babbage del siglo anterior, los británicos no fueron capaces de desarrollar esta notable innovación (Bird 1994). En Estados Unidos, Eckert y Mauchly tuvieron que hacer frente a un grado de escepticismo parecido cuando plantearon la fabricación de ordenadores con fines comerciales. Al final lograron su objetivo, aunque perdieron su independencia por el camino. Se trataba de un escepticismo justificado, si tenemos en cuenta los problemas de ingeniería que había para conseguir que el equipo funcionase debidamente. Sin embargo, hacia mediados de la década de 1950 Eckert y Mauchly consiguieron presentar un ordenador comercial de gran tamaño

llamado UNIVAC, que tuvo una buena acogida por parte de los veinte clientes que lo compraron.

Otras empresas, grandes y pequeñas, también entraron en el negocio de los ordenadores durante esa década, pero a finales de la misma IBM se había colocado claramente a la cabeza. Ello se debió en gran medida a su magnífico departamento de ventas, que se aseguraba de que sus clientes vieran compensada con resultados útiles la gran inversión que habían hecho en equipo electrónico. IBM ofrecía una línea de ordenadores electrónicos diferente para sus clientes empresariales y científicos, así como una línea, que tuvo mucho éxito, de ordenadores pequeños y económicos, como el 1401. Hacia 1960 el transistor, que se inventó en la década de 1940, funcionaba lo suficientemente bien como para reemplazar a los frágiles tubos de vacío de la etapa anterior. La memoria de los ordenadores ahora consistía en una jerarquía de núcleos magnéticos, tambores o discos más lentos y, por último, una cinta magnética de gran capacidad. Para introducir información o programas en estas *macrocomputadoras* todavía había que usar tarjetas perforadas, con lo que se aseguraba la continuidad con el equipo de Hollerith, que era la base de IBM.

En 1964, IBM unificó sus líneas de productos con su *System/360*, que no sólo abarcaba la gama completa de aplicaciones relativas a la ciencia y los negocios (de ahí su nombre), sino que también se presentó como una familia de ordenadores cada vez más grandes, cada uno de los cuales tenía capacidad para ejecutar el *software*

creado para los modelos inferiores. Esto constituyó un paso decisivo que volvió a transformar el sector, como lo había hecho UNIVAC diez años antes. Con ello se reconocía que el *software*, que empezó como una idea de último momento y en la periferia del diseño del soporte físico, se estaba convirtiendo cada vez más en el motor que impulsaba los avances informáticos.

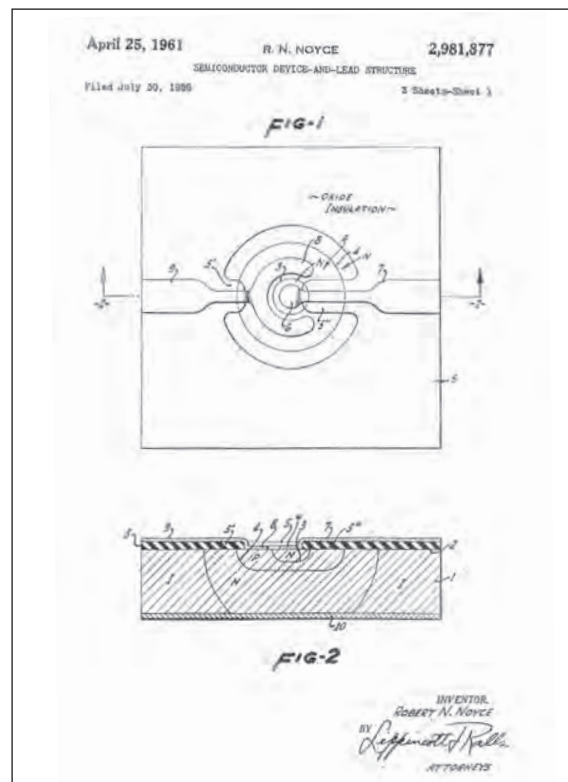
Detrás de IBM en el mercado comercial estaban *los siete enanitos*: Burroughs, UNIVAC, National Cash Register, Honeywell, General Electric, Control Data Corporation y RCA. En Inglaterra, donde en la década de 1940 estuvieron en funcionamiento los primeros ordenadores que incorporaban programas en su memoria, también se desarrollaron productos comerciales, al igual que en Francia. Honrad Zuse, cuyo Z3 ya funcionaba en 1941, también fundó una empresa, quizás la primera del mundo dedicada por entero a la fabricación y venta de ordenadores. Pero, salvo mínimas excepciones, las ventas en Europa nunca se acercaron a las de las empresas estadounidenses. Los soviéticos, aunque competían con Estados Unidos en la exploración espacial, no pudieron hacer lo mismo con los ordenadores. Tuvieron que contentarse con copiar la IBM System/360, con lo que al menos podían aprovechar el *software* que otros habían creado. El motivo por el que la URSS se quedó a la zaga, dada su excelencia técnica y sobre todo matemática, es un misterio. Quizás los encargados de planifi-

cación soviéticos vieron en los ordenadores un arma de doble filo; por un lado facilitarían la planificación estatal, pero por otro harían posible que se compartiera información de manera descentralizada. Desde luego, la falta de una economía de mercado energética, que constituyó un impulsó para los adelantos técnicos de UNIVAC e IBM, fue un factor a tener en cuenta. En cualquier caso, las fuerzas del mercado de Estados Unidos se vieron impulsadas por las enormes sumas de dinero aportadas por el Departamento de Defensa, que subvencionaba la investigación informática para las llamadas operaciones de *control y mando*, así como para la logística y los sistemas de navegación de misiles de a bordo.

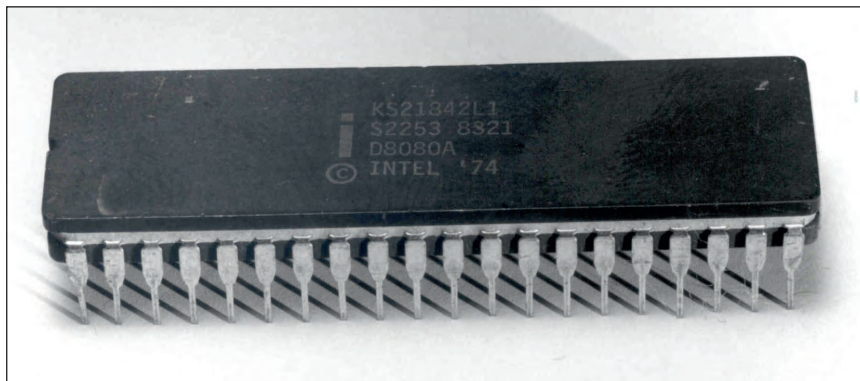
El miniordenador y el chip

Si las tecnologías de la información se hubieran quedado en el punto en el estaban mediada la década de 1960, también ahora estaríamos hablando de una *revolución informática*, tal ha sido el impacto que ha tenido en la sociedad. Pero la tecnología no se quedó quieta; siguió avanzado a un ritmo cada vez más veloz. Pasaron diez años antes de que el transistor saliera de los laboratorios y se empezara a usar de manera comercial y práctica en los ordenadores. Ello tuvo consecuencias para los sistemas de las enormes macrocomputadoras ya mencionados, pero repercutió aún más en los sistemas pequeños. Hacia 1965 hicieron su aparición varios productos nuevos que ofrecían alta velocidad de procesamiento, solidez, un tamaño pequeño y un precio económico, lo que abrió mercados completamente nuevos. El PDP-8, que lanzó aquel año una empresa llamada Digital Equipment Corporation, inauguró esta clase de *miniordenadores*. A partir de aquí surgió un núcleo de fabricantes de miniordenadores en las afueras de Boston. Tanto en lo que se refiere a las personas como a la tecnología, el sector de los miniordenadores es descendiente directo del Proyecto Whirlwind del MIT que subvencionó el Departamento de Defensa (Ceruzzi 1998).

Cuando los diseñadores de ordenadores empezaron a usar los transistores tuvieron que enfrentarse a un problema técnico que en años anteriores había quedado disimulado por la fragilidad de los tubos de vacío. Se trataba de la dificultad que suponía ensamblar, cablear y probar circuitos con miles de componentes diferenciados: transistores, resistencias eléctricas y condensadores. Entre las muchas soluciones que se propusieron a este problema de interconexión estuvieron la de Jack Kilby, de Texas Instruments, y la de Robert Noyce, de Fairchild Semiconductor, cada uno de los cuales registró su patente por separado en 1959. Su invento dio en conocerse con el nombre de *circuito integrado*. Al poder seguir el ejemplo de los pasos que se habían dado con los transistores de silicio, estas empresas lograron comercializar su invento rápidamente: hacia finales de la década de 1960 el chip de silicio se había convertido en el principal dispositivo en los proce-



Robert Noyce, patente para el circuito integrado (chip), 1961. A Noyce, que en aquel momento trabajaba en Fairchild Semiconductor, y a Jack Kilby, que trabajaba en Texas Instruments, se les considera los inventores del circuito integrado. Oficina de Patentes y Marcas de Estados Unidos.



Microprocesador Intel 8080, 1974. El Intel 8080 se utilizó en los primeros ordenadores personales. No fue el primer microprocesador, pero sí el primero que reunió en un solo chip la potencia necesaria para un ordenador de uso práctico. Smithsonian Institution.

sadores de los ordenadores y también había empezado a sustituir a los núcleos de memoria.

Además de inventar con Kilby el circuito integrado, Noyce hizo algo que determinó el rumbo de la ciencia informática. En 1968 abandonó Fairchild y fundó una nueva empresa, llamada Intel, dedicada a la fabricación de chips de memoria como sustitutos de los núcleos magnéticos. El valle de Santa Clara, en la península situada al sur de San Francisco, ya era un centro de microelectrónica, pero el que Noyce fundase allí Intel hizo que su actividad aumentase vertiginosamente. En 1971 un periodista llamó a esta región Silicon Valley: un nombre que hace referencia no sólo a la ingeniería informática que se desarrolla allí sino también a la cultura emprendedora y libre que lo impulsa (Ceruzzi 1998).

Hacia mediados de la década de 1970 la hegemonía de IBM en el mundo de la informática se vio amenazada desde tres frentes. Desde Silicon Valley y las afueras de Boston llegaban noticias de la existencia de sistemas pequeños, pero con una capacidad de procesamiento cada vez mayor. Del Departamento de Justicia de Estados Unidos llegó una demanda antimonopolio, presentada en 1969, en la que se acusaba a IBM de control indebido del sector. Por último, de los ingenieros informáticos que investigaban sobre *software* surgió la noción del uso interactivo de los ordenadores mediante un procedimiento conocido como *tiempo compartido*, que daba a varios usuarios simultáneos la impresión de que aquel ordenador grande y costoso era su máquina de uso personal. El tiempo compartido proporcionaba otra forma de poner capacidad de procesamiento en manos de nuevos grupos de usuarios, pero la promesa del ordenador de uso general económico, similar a la rejilla que suministra electricidad en nuestros hogares, no llegó a materializarse.

Un factor importante de este cambio hacia la informática interactiva fue la creación, en 1964, del lenguaje de programación BASIC en el Dartmouth College del estado de New Hampshire, donde los estudiantes de humanidades, ciencias o ingenierías técnicas descubrieron que sus ordenadores eran más accesibles que los de otras faculta-

des, en los que tenían que presentar sus programas en forma de lote de tarjetas perforadas, codificadas en lenguajes más complicados y esperar a que les llegara el turno.

El ordenador personal

Las diversas críticas al método de cálculo de las macrocomputadoras convergieron en 1975, cuando una empresa poco conocida de Nuevo México sacó al mercado el Altair, que se anunció como el primer equipo informático que costaba menos de 400 dólares. Este equipo apenas se podía llamar *ordenador* y había que añadirle muchos más componentes para conseguir un sistema de uso práctico (Kidwell y Ceruzzi 1994). Sin embargo, el anuncio de Altair desencadenó una explosión de energía creativa que para 1977 había producido sistemas capaces de ejecutar tareas útiles y que empleaban chips de silicio avanzados tanto para el procesador como para la memoria, un disquete (inventado en IBM) para la memoria de masa, y el lenguaje de programación BASIC para permitir que los usuarios escribiesen sus propias aplicaciones de *software*. Esta versión de BASIC se debe a un pequeño equipo dirigido por Bill Gates, quien había dejado sus estudios en Harvard y se había trasladado a Nuevo México para desarrollar *software* para Altair. Con ello se logró arrebatar a IBM la hegemonía sobre el sector informático. Sin embargo, a ninguno de los gigantes que se enfrentaron a IBM les fue particularmente bien durante la siguiente década. Incluso, a principios de los noventa, la Digital Equipment Corporation, a quien debemos en gran medida la existencia del ordenador personal, estuvo a punto de quebrar.

Los ordenadores personales tenían un precio considerablemente más económico, si bien máquinas como las Altair no resultaban apropiadas para nadie que no estuviera muy versado en la electrónica digital y aritmética binaria. En 1977 aparecieron en el mercado varios productos de los que se aseguraba que eran tan fáciles de instalar y usar como cualquier otro electrodoméstico. El más popular fue el Apple II, cuyos fundadores, Steve Jobs y Steve Wozniak, eran el equivalente de Eckert y Maunchy en Silicon Valley: uno era un ingeniero de primera, el otro un visionario que intuyó el potencial de estas máquinas si se hacían accesibles para el gran mercado (Rose 1989). En 1979 apareció un programa llamado *Visicalc* para el Apple II: manejaba filas y columnas de cifras que los contables conocían como *hojas de cálculo*, sólo que con mayor rapidez y facilidad de lo que nadie jamás hubiera imaginado. Una persona que tuviera el *Visicalc* y el Apple II podía ahora hacer cosas que no resultaban fáciles ni para una macrocomputadora. Por fin, tras décadas de promesas, el *software*, es decir, los programas que hacen que los ordenadores hagan lo que uno quiera, pasaron a primer plano, el lugar que en justicia les correspondía. Una década después serían las empresas de *software*, como Microsoft de Bill Gates, las que dominarían las noticias sobre los adelantos de la informática.

A pesar de su reputación de lenta y burocrática, IBM reaccionó con rapidez al reto de Apple y sacó al mercado

su propio PC en 1981. Este PC disponía de una arquitectura abierta que hacía posible a otras empresas suministrar *software*, equipo periférico y tarjetas de circuitos conectables, algo que se alejaba por completo de su filosofía tradicional, aunque muy común en el sector de los minordenadores y otros ordenadores personales. Esta máquina tuvo un éxito comercial mayor del esperado, pues el nombre de IBM daba credibilidad al producto. Empleaba un procesador avanzado de Intel que le permitía tener acceso a mucha más memoria que la competencia. El sistema operativo lo suministró Microsoft y además se puso a la venta un programa de hoja de cálculo, el Lotus 1-2-3, para este PC y los aparatos compatibles con él.

Apple compitió con IBM en 1984 con su Macintosh, con el que sacó de los laboratorios el concepto de *interfaz del usuario* y lo puso al alcance del público en general. La metáfora de ver archivos en la pantalla como una serie de ventanas que se superponen, a las que el usuario accede con un puntero llamado *ratón* se había aplicado por primera vez en la década de 1960 en laboratorios subvencionados por el ejército. A principios de los setenta, un equipo de brillantes investigadores en un laboratorio de Silicon Valley de la Xerox Corporation perfeccionó este concepto. Pero fue Apple quien lo convirtió en un éxito comercial; Microsoft le siguió con su propio sistema operativo, Windows, que se lanzó casi coincidiendo con el Macintosh, pero que no se comercializó con éxito hasta 1990. A lo largo de la siguiente década prosiguió la batalla entre la arquitectura de Apple y la promovida por IBM, que utilizaba procesadores de Intel y un sistema de *software* de Microsoft.

Las primeras conexiones de red

Durante la década de 1980 los ordenadores personales acercaron la informática a los ciudadanos. Muchos individuos los utilizaban en el trabajo, y unos cuantos también tenían uno en casa. La tecnología, aunque todavía algo

desconcertante, había dejado de ser un misterio. Ahora bien, aunque los ordenadores personales dominaban la prensa diaria, las respetadas macrocomputadoras seguían dominando la industria por lo que se refería al valor en dólares del equipo y del *software* que incorporaban. Aunque no podían competir con las aplicaciones de los programas para PC tales como las hojas de cálculo y los procesadores de texto, sí eran necesarias para las operaciones que requerían manejar grandes cantidades de datos. A principios de la década de 1970 estos ordenadores empezaron a cambiar las tarjetas perforadas por operaciones interactivas realizadas con el teclado y otros terminales que tenían el mismo aspecto físico que el de un ordenador personal. Los grandes sistemas de bases de datos en línea se convirtieron en algo habitual y poco a poco empezaron a transformar las actividades comerciales y gubernamentales de los países industrializados. Entre las aplicaciones más visibles están los sistemas de reservas aéreas, los de información al cliente y de facturación para las empresas de servicios públicos y compañías de seguros, así como los inventarios informatizados para minoristas. La combinación de sistemas de bases de datos y de facturación en línea, de números de teléfono gratuitos, de verificación de tarjetas de crédito y facturación telefónica transformó a la humilde rama minorista de venta por correo en una de las grandes fuerzas de la economía estadounidense.

Para todas estas actividades se necesitaban macrocomputadoras grandes y costosas y que dispusieran de un *software* diseñado a medida, lo cual suponía un enorme gasto para el cliente. Existía la tentación de conectar una serie de ordenadores personales baratos que ejecutasen paquetes de *software* económicos y de bajo mantenimiento, pero esto no era viable. Puede que si se engancha un grupo de caballos a un carro se ayude a arrastrar más peso, pero no hará que el carro vaya más rápido. Y hasta esto tiene sus limitaciones, pues al carretero cada vez le resultará más difícil que todos los caballos tiren en la misma dirección. El problema con la informática era parecido y quedó expresado en la *Ley de Grosch*: por el mismo dinero, rinde más el trabajo que realiza un ordenador grande que dos pequeños (Grosch 1991).

Pero esto iba a cambiar. En el Centro de investigación de Palo Alto de Xerox en 1973, donde se habían logrado tantos avances relacionados con la interfaz de usuario, se inventó un método de conexión de redes que dejó la ley de Grosch obsoleta. Sus creadores la llamaron *Ethernet*, en honor al medio (éter) que, según los físicos del siglo XIX, transportaba la luz. Ethernet hizo posible conectar entre sí los ordenadores pequeños de una oficina o edificio, y con ello compartir la memoria de masa, las impresoras láser (otro invento de Xerox) y que los usuarios de los ordenadores intercambiaran mensajes de correo electrónico. Al tiempo que Ethernet hacía posible la conexión de redes local, un proyecto financiado por la Agencia de Proyectos de Investigación Avanzada en Defensa (ARPA) hacía lo propio para conectar



Ordenador Personal Altair, 1974. Una pequeña empresa de aficionados a la informática, MITS, de Albuquerque, Nuevo México, presentó en 1974 este ordenador, que fue una revolución en el terreno de los ordenadores personales. Smithsonian Institution.



Estación de trabajo Alto de Xerox, 1973. La Alto se diseñó y construyó en el Centro de investigación Palo Alto de Xerox (Palo Alto Research Center, PARC) de California. Fue precursora del uso del ratón y de la interfaz gráfica del usuario, que con el tiempo pasaría a ser habitual en los ordenadores. Los Altos estaban conectadas entre sí por medio de Ethernet y a su vez con impresoras láser. Ambas son innovaciones de Xerox. Smithsonian Institution.

ordenadores geográficamente dispersos. Tenía como objeto que las comunicaciones militares se mantuvieran seguras en caso de guerra, cuando los tramos de una red podían ser destruidos. Las primeras redes militares que provenían del Proyecto Whirlwind tenían unidades de mando central, y por ello era posible atacar al centro de control de la red. Estas unidades se encontraban en edificios sin ventanas, reforzados con estructuras de hormigón, pero si sufrían daños la red dejaba de funcionar (Abbate 1999).

ARPA financió la labor de un grupo de investigadores que desarrollaron una alternativa en la que se dividió la información en *paquetes*, cada uno los cuales recibía la dirección de un ordenador receptor y circulaban a través de la red de ordenadores. Si uno o más ordenadores en la red no funcionaban, el sistema encontraría otra ruta. El receptor reunía los paquetes y los convertía en una copia fiel del documento original que había transmitido. Hacia 1971 ARPANET contaba con quince nodos de conmutación repartidos por todo el país y en los nueve años siguientes creció con gran rapidez. En un principio tenía como objeto enviar conjuntos de datos grandes o programas de un nodo a otro, pero poco después de que la red entrase en funcionamiento la gente empezó a utilizarla para intercambiar mensajes breves. En un primer momento se trataba de un proceso laborioso, pero en 1973 Ray Tomlinson, un ingeniero de la empresa Bolt Beranek and Newman

de Cambridge, Massachussets, hizo que esto cambiase. A Tomlinson se le ocurrió la sencilla idea de separar el nombre del receptor del mensaje y el de su ordenador con el símbolo @, uno de los pocos símbolos no alfabéticos de los que disponía el panel de mandos del teletipo que ARPANET empleaba en aquella época. Y así es como se concibió el correo electrónico, y con él, el símbolo de la era de las conexiones de red.

La presión ejercida para que ARPANET se pudiera destinar al envío de correos electrónicos y a otros usos que no fueran militares fue tan grande que la red terminó por escindirise. Una parte quedó bajo el control militar; la otra se cedió a la National Science Foundation (NSF), un organismo civil financiado por el Estado que subvencionó proyectos de investigación no sólo para ampliar esta red, sino también para hacer que se interconectasen los diferentes tipos de redes (por ejemplo, las que utilizaban radios en lugar de cables). Los investigadores empezaron a llamar al resultado de todo ello *Internet*, para reflejar así su naturaleza heterogénea. En 1983 las redes adoptaron un conjunto de normas para la transmisión de datos con esta interconexión llamado *Protocolo de Control de Transmisión/Protocolo de Internet* (Transmission Control Protocol/Internet Protocol, TCP/IP). Estos protocolos se siguen usando en la actualidad y constituyen la base de la Internet actual (Aspray y Ceruzzi 2008).

Estas redes de conexión local y remota encajaron a la perfección con otros cambios que se estaban desarrollando en el *software* y el *hardware* de los ordenadores. Salió un nuevo tipo de ordenador denominado *estación de trabajo*, que a diferencia de los PC se adecuaba mejor a las conexiones de redes. Otra diferencia fundamental es que utilizaba un sistema operativo llamado *UNIX*, que si bien era de difícil manejo para el consumidor, se ajustaba muy bien a las conexiones de red y a otros programas avanzados. UNIX fue creado por los laboratorios Bell, la sección dedicada a la investigación del monopolio de telefonía AT&T que regula el gobierno estadounidense. Los grupos de estaciones de trabajo, conectados entre ellos localmente por Ethernet, y por Internet a grupos de terminales similares por todo el mundo, por fin suponían una alternativa real a las grandes instalaciones de macrocomputadoras.

La era de Internet

La National Science Foundation (NSF), una agencia gubernamental estadounidense, no podía permitir el uso comercial de la parte de Internet que estaba bajo su control. Sí podía, sin embargo, ceder el uso de los protocolos de Internet a cualquiera que quisiera utilizarlos por muy poco dinero o de forma gratuita, a diferencia de lo que ofrecían empresas de ordenadores como IBM. Con el aumento de usuarios de Internet, la NSF se vio presionada para ceder su gestión a empresas comerciales. En 1992 el Congreso de Estados Unidos aprobó una ley con la que terminó de hecho la prohibición de su uso comercial, por lo que se puede

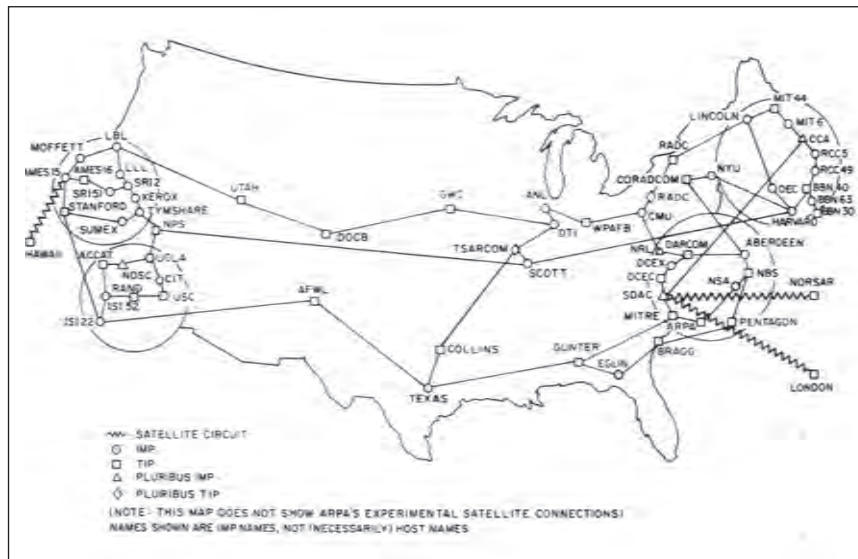
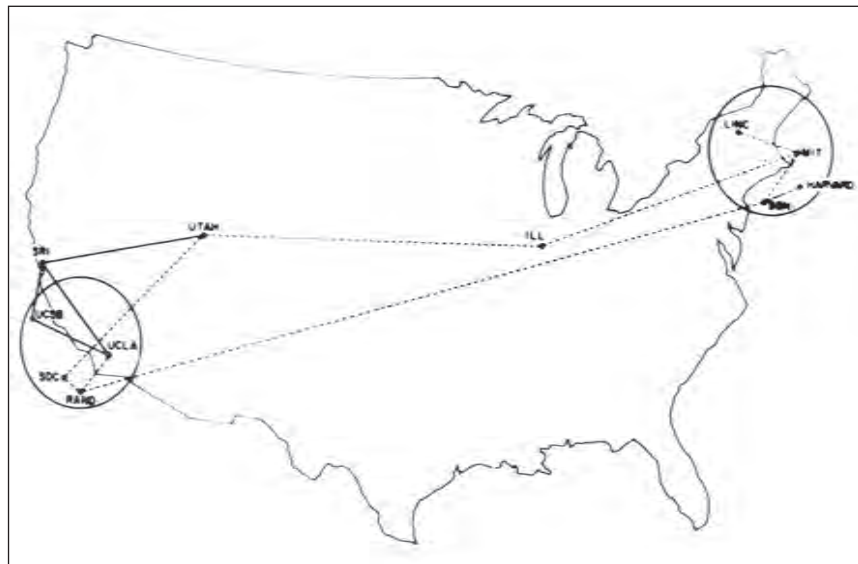
decir que la aprobación de esta ley marcó el comienzo de la era de Internet. Ahora bien, esto no es completamente cierto, pues el gobierno estadounidense retuvo el control sobre el plan de direcciones de Internet, por ejemplo, los sufijos *.com*, *.edu*, etc., que permiten a los ordenadores saber adónde se envía un mensaje electrónico. A principios del siglo XXI, una serie de países pidió que dicho control pasara a la Organización de las Naciones Unidas, pero hasta ahora Estados Unidos se ha mostrado reacio. Se trata realmente de un recurso que se ofrece a todos los países del mundo, pero el registro maestro de los nombres de dominio lo gestiona una empresa privada estadounidense a la que el Departamento de Comercio concede esta autoridad.

Esta actividad política se vio complementada por adelantos significativos en la tecnología informática, lo que

supuso un nuevo impulso para la difusión de Internet. Para 1990 las costosas estaciones de trabajo de UNIX habían cedido el paso a los ordenadores personales que utilizaban procesadores avanzados, en especial un procesador llamado *Pentium*, que suministraba Intel. En lo que respecta al *software*, salieron versiones nuevas del sistema operativo Windows de Microsoft en las que venían instalados los protocolos de Internet y otros programas de conexión de redes. Esta combinación proporcionó a los PC una potencia equivalente a la de las estaciones de trabajo. Es raro encontrar UNIX en un PC, aunque los servidores de mayor potencia y los denominados *routers* que realizan las conmutaciones básicas de Internet lo siguen usando. Una variante de UNIX llamada *Linux*, creada en 1991 por Linus Torvalds en Finlandia, se presentó como una alternativa gratuita o muy barata al sistema Windows de Microsoft, y tanto éste como el *software* relacionado con él lograron hacerse con una cuota de mercado pequeña, si bien significativa. Estos programas pasaron a conocerse como *software de código abierto*, el cual se define como *libre*, pero no sin *restricciones* (Williams 2002).

Mientras esta actividad se desarrollaba en los laboratorios gubernamentales y universitarios, los usuarios de PC empezaban a descubrir las ventajas de las conexiones de red. Los primeros ordenadores personales como el Apple II no tenían una gran capacidad para conectarse a una red, pero aficionados con mucha imaginación consiguieron desarrollar formas ingeniosas de comunicarse. Utilizaron un dispositivo llamado *modem* (modulador-demodulador) para transmitir datos informáticos lentamente a través de las líneas telefónicas. En esta empresa se vieron asistidos por una decisión tomada por el monopolio de telefonía estadounidense, según la cual los datos que se enviaban por líneas telefónicas recibirían la misma consideración que las llamadas de voz. Las llamadas locales eran, de hecho, gratuitas en Estados Unidos, pero las llamadas a larga distancia resultaban caras. Estos entusiastas de los ordenadores personales encontraron formas de reunir mensajes localmente y luego enviarlos de un lado a otro del país por la noche, cuando las tarifas eran más baratas (esto dio lugar a *FidoNet*, llamada así por un perro que iba a buscar información, como los perros cuando corren a buscar un objeto que se ha lanzado). También surgieron empresas comerciales para abastecer este mercado; alquilaban números de teléfono en las áreas metropolitanas y cobraban una tarifa a los usuarios por conectarse. Uno de los más importantes fue *The Source*, que se fundó en 1979 y que tras atravesar un periodo de dificultades financieras se reorganizó y convirtió en la base para America Online, el servicio de conexiones de red personal más popular desde la década de 1980 hasta finales de la de 1990.

Estos sistemas comerciales y personales son importantes porque con ellos las conexiones de redes cobraron una dimensión social. ARPANET era una red militar, y sus responsables desaprobaban su uso frívolo y comercial. Pero



ARPANET, 1970 y 1974. La actual Internet procede de esta red subvencionada por el ejército, que creció con gran rapidez desde sus inicios, en la década de 1970. Agencia de investigación de proyectos avanzados de defensa, Departamento de Defensa de Estados Unidos.



Teletipo Model ASR-33. La ARPANET utilizó este teletipo modificado como un terminal. Nótese el símbolo @, que se adoptó para el correo electrónico y se ha convertido en un icono de las conexiones de red. Smithsonian Institution.

las redes personales, como los teléfonos de particulares a través de los que se transmitían estos mensajes, se utilizaron desde el principio para *chats*, debates informales, noticias y servicios comerciales. Una de las redes comerciales, Prodigy, también incluía gráficos a color, otro de los elementos básicos de la Internet de hoy. Las historias sobre Internet que hacen subrayan la importancia de ARPANET están en lo correcto: ARPANET fue su predecesora técnica, y sus protocolos surgieron de la labor de investigación del ARPA. Sin embargo, para que una historia de Internet sea completa, también hay que tener en cuenta su dimensión social y cultural, la cual surgió a partir de Prodigy, AOL, así como de la comunidad de usuarios aficionados.

Hacia finales de la década de 1980 era evidente que las redes de ordenadores resultaban ventajosas para hogares y oficinas. No obstante, la red que se estaba creando con el apoyo de la National Science Foundation, era una de las muchas aspirantes. Los informes comerciales de aquellos años defendían un tipo de red completamente diferente, me refiero en concreto a la ampliación de la televisión por cable hasta alcanzar una multitud de canales nuevos, quinientos, según un pronóstico generalizado del momento. Esta nueva configuración de la televisión permitiría cierto grado de interactividad, pero ello no sería posible con un ordenador personal de uso domésti-

co. Se trataba de un producto lógico de los objetivos de marketing de los sectores de televisión y entretenimiento. Entre la comunidad de científicos y profesionales informáticos, las conexiones de red vendrían dadas a través de un conjunto bien estructurado de protocolos llamado *interconexión de sistema abierto* (Open Systems Interconnection, OSI), que reemplazaría a Internet, de estructura más abierta. Nada de esto ocurrió, en gran manera debido a que Internet, a diferencia de los otros proyectos, se diseñó para permitir el acceso a redes diferentes sin estar vinculada a un monopolio regulado por el gobierno, grupo empresarial privado o sector en particular. Hacia mediados de la década de 1990 las redes privadas como AOL establecieron conexiones con Internet y los protocolos OSI cayeron en desuso. Paradójicamente, porque Internet era de acceso gratuito y no había sido concebida para un uso comercial determinado, pudo convertirse en la base de tanta actividad comercial una vez que salió del control del gobierno de Estados Unidos, después de 1993 (Aspray y Ceruzzi 2008).

En el verano de 1991 investigadores del Laboratorio Europeo de Física de Partículas CERN sacaron un programa llamado World Wide Web. Consistía en un conjunto de protocolos que operaban por encima de los protocolos de Internet y permitían un acceso muy flexible y generalizado



Red de la National Science Foundation (NSF), 1991. La NSF apoyó la transición de las conexiones de red del uso militar al civil. Como agencia gubernamental, siguió restringiendo su uso a los ámbitos de la educación y la investigación. Una vez desaparecieron estas restricciones, poco tiempo después de que se hiciera este mapa, nació la Internet comercial tal y como hoy la conocemos. U.S. National Science Foundation.

a la información almacenada en la red en diversos formatos. Al igual que ocurrió con Internet, esta característica de acceso a todo tipo de formatos, máquinas, sistemas operativos y normas fue lo que hizo que su uso se generalizase rápidamente. En la actualidad y para la mayor parte de los usuarios, World Wide Web e Internet son sinónimos; ahora bien, es más apropiado decir que esta última constituyó la base de la primera. El principal creador de la World Wide Web fue Tim Berners-Lee, que en aquella época trabajaba en el CERN. Según recuerda, lo que le inspiró su creación fue ver cómo físicos de todo el mundo se reunían para debatir cuestiones científicas en los edificios del CERN. Además de la World Wide Web, Berners-Lee también desarrolló otro programa mediante el que se facilitaba el acceso a ésta desde un ordenador personal. Este programa, denominado *buscador*, fue un factor clave adicional en la popularización del uso de Internet (Berners-Lee 1999). Su buscador tuvo sólo un uso limitado y fue rápidamente reemplazado por uno más sofisticado llamado *Mosaic*, que se creó en 1993 en la Universidad de Illinois, en Estados Unidos. Al cabo de dos años los principales creadores de Mosaic abandonaron Illinois y se trasladaron a Silicon Valley en California, donde fundaron una empresa que se llamó Netscape. Los usuarios particulares podían descargar su buscador, *Navigator*, de manera gratuita, pero los comerciales tenían que pagar. El éxito casi instantáneo de Netscape supuso el comienzo de la *burbuja* de Internet, en virtud de la cual cualquier valor que estuviese remotamente relacionado con ella cotizaba a unos precios desorbitados. Mosaic desapareció, pero Microsoft compró sus derechos y lo convirtió en la base de su propio buscador, Internet Explorer, que en la actualidad es el medio más utilizado de acceso a la Web y a Internet en general (Clark 1999).

Conclusión

La historia de la informática empezó de manera lenta y metódica, y luego se disparó con la llegada de las conexiones de red, los buscadores y, ahora, con los dispositivos portátiles. Todo intento por trazar su trayectoria reciente está condenado al fracaso. Esta fuerza que la impulsa viene definida en la Ley de Moore, uno de los fundadores de Intel, según la cual los chips de silicio duplican su capacidad cada dieciocho meses (Moore 1965). Esto es lo que lleva ocurriendo desde 1960, y, a pesar de que periódicamente se pronostica que esto terminará pronto, parece que aún no es el caso. Asimismo, la capacidad de la memoria de masa, en especial de los discos magnéticos, y de la anchura de banda de los cables de telecomunicaciones y otros canales ha ido aumentando a un ritmo exponencial. Todo ello hace que los ingenieros estén atrapados en una rutina de la que no tienen escapatoria: cuando les piden que diseñen un producto no lo hacen pensando en la capacidad de los chips que hay en ese momento, sino en la potencia que calculan que tendrán cuando el producto salga a la venta, lo cual, a su vez, obliga a los fabricantes de chips a sacar uno que satisfaga esas expectativas. En la prensa general y especializada siempre se pueden encontrar predicciones en las que se indica que esto algún día se acabará: al menos cuando los límites de la física cuántica hagan imposible diseñar chips con mayor densidad. Sin embargo, a pesar de todos estos pronósticos que señalan que la Ley de Moore llegará a su fin, todavía no ha ocurrido, y mientras siga siendo válida es imposible predecir qué *camino* seguirá la informática, incluso, el año que viene. Pero esto es lo que convierte a esta era en una de las más emocionantes de la historia, siempre y cuando uno sea capaz de sobrellevar la velocidad a la que se producen los cambios tecnológicos.

Bibliografía

- Abbate, J. *Inventing the Internet*. Cambridge, Massachusetts: MIT Press, 1999.
- Aspray, W., ed. *Computing Before Computers*. Ames, Iowa: Iowa State University Press, 1990.
- , y P. E. Ceruzzi, eds. *The Internet and American Business*. Cambridge, Massachusetts, 2008.
- Berners-Lee, T. y M. Fischetti. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*. San Francisco: Harper, 1999.
- Bird, P. *LEO: The First Business Computer*. Berkshire, Reino Unido: Hasler Publishing, 1994.
- Burks, A. R. y W. Arthur. *The First Electronic Computer: The Atanasoff Story*. Ann Arbor, Michigan: University of Michigan Press, 1988.
- Ceruzzi, P. E. *Reckoners: the Prehistory of the Digital Computer, From Relays to the Stored Program Concept, 1935-1945*. Westport, Connecticut: Greenwood Press, 1983.
- , *A History of Modern Computing*. Cambridge, Massachusetts: MIT Press, 1998.
- Clark, J. y O. Edwards. *Netscape Time: The Making of the Billion-Dollar Start-Up that Took on Microsoft*. Nueva York: St. Martin's Press, 1999.
- Eames, Ch. y R. Offices of. *A Computer Perspective: Background to the Computer Age*. Cambridge, Massachusetts: Harvard University Press, 1990.
- Eckert, W. J. *Punched Card Methods in Scientific Calculation*. Nueva York: IBM Corporation, 1940.
- Grosch, H. R. J. *Computer: Bit Slices from a Life*. Novato, California: Third Millennium Books, 1991.
- Kidwell, P. A., y P. E. Ceruzzi. *Landmarks in Digital Computing: A Smithsonian Pictorial History*. Washington, D. C.: Smithsonian Institution Press, 1994.
- Merzbach, U. *Georg Scheutz and the First Printing Calculator*. Washington, D. C.: Smithsonian Institution Press, 1977.
- Moore, G. E. «Cramming More Components onto Integrated Circuits», *Electronics*, (19 de abril de 1965): 114-117.
- Randall, B., ed. *The Origins of Digital Computers: Selected Papers*. Berlín, Heidelberg y Nueva York: Springer-Verlag, 1975.
- Redmond, K. C. y Th. M. Smioth. *Project Whirlwind: The History of a Pioneer Computer*. Bedford, Massachusetts: Digital Press, 1980.
- Rose, F. *West of Eden: The End of Innocence at Apple Computer*. Nueva York: Penguin Books, 1989.
- Stern, N. *From ENIAC to UNIVAC: An Appraisal of the Eckert-Mauchly Computers*. Bedford, Massachusetts: Digital Press, 1981.
- Williams, S. *Free as in Freedom: Richard Stallman's Crusade for Free Software*. Sebastopol, California: O'Reilly, 2002.

los ordenadores y la exploración espacial

PAUL E. CERUZZI

La puesta en órbita por parte de la Unión Soviética de dos satélites Sputnik en el otoño de 1957 fue una conmoción para muchos estadounidenses. Aunque para los servicios de inteligencia de Estados Unidos no supuso una sorpresa, sí lo fue para los ciudadanos de a pie, y los lanzamientos demostraron, más allá de toda duda, que la Unión Soviética llevaba la delantera a Estados Unidos no sólo en la construcción de satélites, sino también de cualquier cohete impulsor, capaz además de transportar armamento. Entre las reacciones de Estados Unidos al Sputnik estuvo la creación de dos agencias, una de ellas dependiente del Departamento de Defensa, la otra civil. La primera se llamó Advanced Research Projects Agency (Agencia de proyectos de investigación avanzada) o ARPA, recientemente más conocida como DARPA. La misión del ARPA era clara: apoyar investigación a largo plazo que garantizara que Estados Unidos no fuera cogido nunca más desprevenido como lo estaba en el momento del lanzamiento de los Sputnik. Unas de sus áreas de investigación eran los misiles y la exploración espacial; a finales de 1958 el grueso de esta tarea fue transferido a otra agencia, de control civil: la National Air and Space Administration o NASA (Norberg y O'Neil 1996).

En los cincuenta años que siguieron a su creación ambas agencias acumularon un número notable de logros, pero sin duda los más importantes fueron dos: a principios de la década de 1960 DARPA diseñó y puso en funcionamiento una red de ordenadores conocida como ARPANET, que fue la inspiración técnica de lo que hoy es Internet.

Por su parte la NASA, en respuesta a un desafío planteado por John F. Kennedy en 1961, consiguió enviar con éxito entre 1969 y 1972 a una docena de astronautas a la Luna y hacerlos regresar sanos y salvos.

Mediada la década de 1990 Internet pasó rápidamente de ser una red conocida sólo para los informáticos y otros especialistas a una herramienta al alcance del ciudadano medio del mundo industrializado. En Estados Unidos, la cadena de televisión sin ánimo de lucro Public Broadcasting Service produjo un programa que constaba de varios capítulos para documentar el ascenso meteórico de este fenómeno tecnológico. Recibió el caprichoso título de «Nerds 2.0.1.: A Brief History of the Internet» (Pitagorines 2.0.1.: breve historia de Internet) (Segaller 1998). El título sugería que Internet era la creación de estos pitagorines o locos por los ordenadores, en su mayoría hombres jóvenes, pocos de ellos de más de treinta años, cuya obsesión por los ordenadores dio lugar a este gigantesco fenómeno que ha cambiado nuestra realidad cotidiana. En casi todos los episodios del programa el narrador llamaba la atención sobre el contraste existente entre los logros de las dos agencias fundadas al mismo tiempo: Internet como heredera de la actividad de ARPA y la llegada del hombre a la Luna como resultado del trabajo de la NASA.

El núcleo del programa insistía en esta teoría. Afirmaba —con razón— que Internet descende de ARPANET, una red informática diseñada y patrocinada por el ejército de Estados Unidos. Pero iba más allá, argumentando que los

alunizajes eran tan sólo una hazaña aislada, con escaso o ningún impacto en la sociedad, mientras que Internet era una tecnología revolucionaria que afectaba, y afecta, las vidas de la gente corriente en todo el mundo.

Medio siglo después de la creación de estas dos agencias podemos repasar los logros alcanzados en informática y exploración del espacio y preguntarnos acerca de la relación entre las tecnologías implicadas en cada uno. Tanto en tecnología aeroespacial como en informática ha habido enormes progresos, pero el futuro no ha resultado ser del todo como muchos esperaban.

A finales de la década de 1960 muchos influyentes científicos informáticos predijeron que los ordenadores alcanzarían la «Inteligencia Artificial» (IA) y se convertirían en nuestros criados personales, tal vez incluso en compañeros (McCorduck 1979). Los escritores de ciencia-ficción adoptaron este tema para sus libros y retrataron ordenadores provistos de inteligencia artificial ya fuera como asistentes personales, como en el caso de los robots de la serie *La guerra de las galaxias*, o como nuestros enemigos, como el malévolo HAL en *2001: Una odisea del espacio*. Pero a pesar de lo recurrente del tema, no llegó a suceder en la realidad. La inteligencia artificial continúa siendo una meta científica huidiza. Sin embargo, fuera de los estrechos confines de la comunidad de inteligencia artificial de científicos informáticos, este «fracaso» no parece preocupar a nadie. La razón es simple: la llegada del ordenador personal, de Internet, del teléfono inalámbrico y de otros avances tecnológicos han elevado el mundo de la informática a niveles que superan lo que muchos predijeron cuando el hombre pisó la Luna. No podemos conversar con ellos como lo haríamos con alguien de carne y hueso, pero estos sistemas presentan una cantidad sorprendente de lo que podríamos llamar «inteligencia», que procede más de su capaci-

dad de procesar potencia y memoria que de su propósito intrínseco: ser un sustituto artificial del cerebro humano.

En el ámbito de la exploración espacial, las sucesivas misiones Apollo a la Luna generaron predicciones que nunca se hicieron realidad: estaciones permanentes en aquel planeta, hoteles en órbita, misiones a Marte tripuladas por seres humanos. Nada de esto ha sucedido aún y, sin embargo, los avances en tecnología espacial han sido verdaderamente notables. La Tierra está ahora rodeada de comunicaciones y los satélites meteorológicos forman parte de nuestra vida cotidiana. El Global Position System o GPS, así como sus proyectados equivalentes europeo y asiático, proporcionarán servicios de localización y de estimación de tiempo de desplazamientos precisos y baratos. Las sondas robóticas espaciales han iniciado una exploración de Marte y otros planetas lejanos que rivalizan con los viajes de expedición de cualquier época anterior. Los telescopios espaciales que operan en longitudes de onda visibles y no visibles señalan el inicio de una era que promete ser de las más emocionantes de la historia de la humanidad (Dick y Launius 2007).

En cuanto al campo de la informática, sólo los avances en la capacidad de memoria y poder de procesamiento, además del funcionamiento en red, han compensado con creces la imposibilidad de los ordenadores de pensar como los humanos. En el ámbito de la exploración espacial los avances arriba descritos no han eliminado la frustración por no haber conseguido establecer una presencia humana significativa fuera de nuestro planeta (en el campo de los aviones que vuelan dentro de los límites de la atmósfera terrestre, las últimas décadas han sido testigo de fracasos similares. A finales de la década de 1960 se rompió la barrera del sonido, pero, con excepción de unas pocas naves del ejército, la mayoría de los aviones sigue volando por debajo de esa barrera. Los aviones comerciales siguen volando a aproximadamente la misma velocidad y altitud que los primeros aparatos comerciales que entraron en funcionamiento en los años cincuenta del pasado siglo. El supersónico Concorde, aunque un prodigio de la técnica, fue un fracaso comercial y pronto estuvo fuera de servicio).

De ahí la tesis del programa de televisión: que la casi desapercibida red informática de ARPA hace palidecer los más espectaculares logros en aeronáutica y exploración del espacio de la NASA. Una opinión que muchos espectadores aparentemente compartían, por muchos argumentos en contra que adujeran la NASA y otros forofos del espacio.

Durante los últimos sesenta años la informática y la navegación espacial han estado estrechamente relacionadas y es prácticamente imposible estudiar sus trayectorias por separado. La invención de la computadora electrónica digital, que se produjo en varios lugares distintos entre 1940 y 1950, a menudo estuvo ligada a la resolución de problemas en los campos de la astronomía y la aerodinámica, o concebida para apoyar las tecnologías de diseño y producción de aviones, control de tráfico aéreo, armamen-

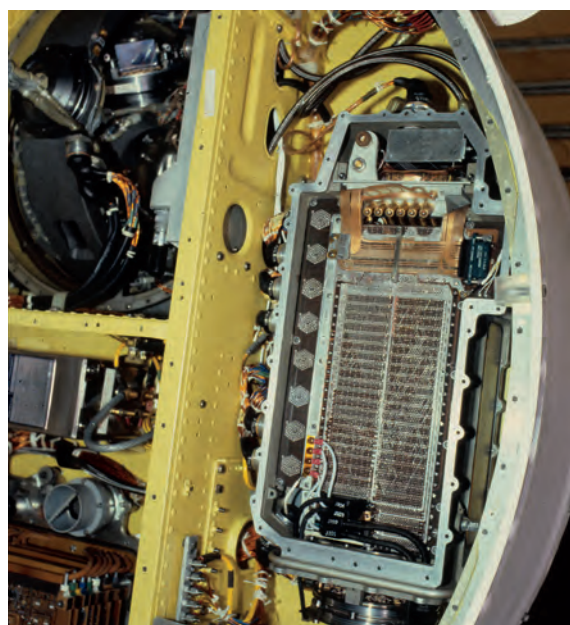


CRAY-1 Supercomputer, c 1976. La CRAY-1, diseñada por Seymour Cray, fue la primera «supercomputadora» capaz de competir con túneles de viento en el análisis de la atmósfera y en el diseño de naves espaciales. CRAY Research, Inc.

to anti nuclear y, más tarde, desarrollo de misiles direccionados. Una de las fuentes de inspiración para la creación de ARPANET fue la necesidad de adaptar las redes de comunicación a la crisis de control de armamento desencadenada por el desarrollo de misiles balísticos y bombarderos a reacción. No se trataba únicamente de diseñar una red capaz de sobrevivir a un ataque nuclear, como afirman muchas historias que circulan por ahí; también respondía a la necesidad de poseer un sistema de comunicaciones que fuera lo suficientemente flexible y resistente como para adaptarse al nuevo entorno espacial que se abrió después de la Segunda Guerra Mundial (Abbate 1999).

A partir de 1945 e iniciada la Guerra Fría contra la URSS, la comunidad aeroespacial estadounidense empezó a disponer de grandes sumas de dinero procedente de los presupuestos de Defensa. Ello impulsó el desarrollo de la informática digital, que progresó en Estados Unidos mucho más que en Inglaterra, donde se habían diseñado las primeras computadoras dedicadas al descifrado de códigos, los primeros programas informáticos y el primer ordenador comercial. Parte de ese dinero se malgastó, pero el apoyo militar de Estados Unidos, dedicado en gran medida aunque no exclusivamente, al desarrollo aeroespacial, fue un poderoso motor para el avance tecnológico.

Por su misma naturaleza, un ordenador digital es un mecanismo de múltiples funciones. Si se puede diseñar un programa adecuado para él —esto es una condición imprescindible— entonces puede usarse para una gran variedad de tareas. Esta cualidad, descrita por primera vez



Sistema de guiado del Minuteman III, c. 1970. El Minuteman, un potente misil balístico desarrollado por las Fuerzas Aéreas estadounidenses a principios de la década de 1960, fue pionero en el uso de componentes electrónicos. Para el primer Minuteman, las Fuerzas Aéreas desarrollaron un programa llamado de «Alta Fiabilidad» para sus componentes electrónicos. El Minuteman III, un modelo posterior, fue pionero en el uso del recientemente inventado circuito integrado. Smithsonian Institution.

en términos teóricos por el matemático inglés Alan Turing en la década de 1930, pronto diferenció al ordenador del resto de las máquinas, generalmente diseñadas y optimizadas para una única función. Así pues la navegación aeroespacial fue tan sólo uno de los campos en los que la informática encontró aplicaciones. La década de 1950 fue testigo de un aumento continuado en la potencia y la capacidad de memoria de los ordenadores, algo que sucedió paralelamente al desarrollo de programas de *software* de uso general, tales como el lenguaje de programación FORTRAN, y de otros específicos empleados para diseño asistido por ordenador (CAD), fabricación asistida por ordenador (CAM), análisis de esfuerzo y dinámica de fluidos.

A diferencia de las aplicaciones de la informática en, digamos, la banca o las finanzas, las de la aviación poseen una limitación añadida. Hasta aproximadamente 1960 los ordenadores eran de gran tamaño, frágiles y consumían grandes cantidades de energía. Eso restringía sus aplicaciones en la aviación al ámbito terrestre: reserva de vuelos, análisis de túneles de viento, CAD, CAM y procedimientos similares. El potencial del ordenador para convertirse en una máquina de uso universal en la industria de la aviación estaba, a juicio de Turing, lastrado por la necesidad de adaptarse a los rigores de la navegación aérea y espacial. La comunidad aeroespacial y militar, que en la década de 1950 en Estados Unidos disponía de enormes recursos económicos, estaba por tanto en condiciones de moldear el curso de la industria informática en sus años clave. A su vez, conforme los ordenadores ganaban en ligereza, manejabilidad y resistencia, la informática empezó a influir en el campo de la navegación espacial a lo largo de una década marcada por los rápidos cambios en la tecnología aérea (Ceruzzi 1989).

El transistor, inventado a finales de los años cuarenta del siglo pasado, fue el primer avance tecnológico que se enfrentó a problemas de tamaño, fiabilidad y peso. Fue necesario un largo periodo de desarrollo, sin embargo, antes de que el transistor con chip de silicio fuera lo suficientemente fiable como para permitir que los ordenadores se convirtieran en aparatos pequeños, resistentes y de bajo consumo. Los ordenadores transistorizados se incorporaron por primera vez a sistemas de direccionamiento de misiles alrededor de 1960. En 1959 dos ingenieros, Jack Kilby, de la compañía Texas Instruments, y Robert Noyce, de Fairchild Instruments, fueron un paso más allá y desarrollaron circuitos en los cuales se situaban en un solo chip de material (primero germanio, después silicio) varios transistores y otros componentes. Había nacido el circuito integrado o chip de silicio. Ni Noyce ni Kilby trabajaban entonces en la industria aeroespacial y, sin embargo, fueron las necesidades de esta industria las que determinaron la invención del chip de silicio. En los doce años que mediaron entre la invención del transistor y la del chip de silicio, las fuerzas aéreas estadounidenses emprendieron una campaña para mejorar la fiabilidad de los circuitos electrónicos en general. Por entonces la aviación estaba desarrollando misiles

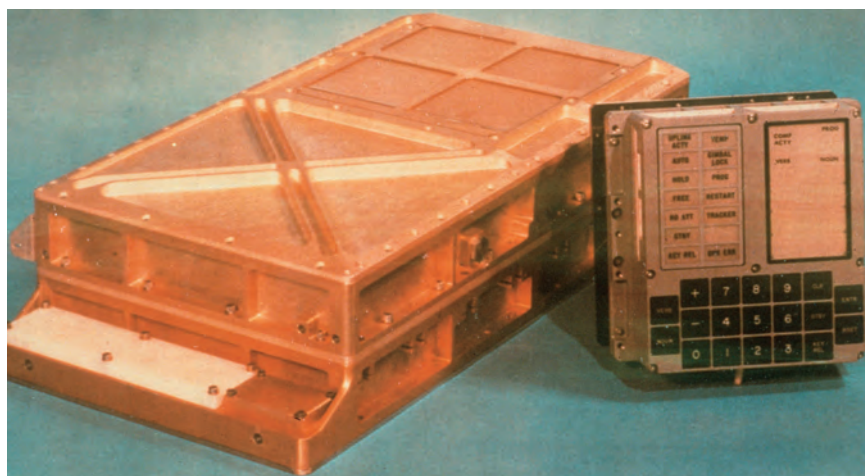
balísticos: armas por valor de un millón de dólares que en ocasiones explotaban en la plataforma de lanzamiento debido al fallo en un componente electrónico de menos de un dólar de coste. La industria electrónica de la década de 1950 basaba sus modelos económicos en un mercado de consumo en el cual los bajos costes de fabricación y no la alta calidad era lo que garantizaba los beneficios. Los consumidores de la época simplemente aceptaban el fallo ocasional de alguno de los componentes, del mismo modo que hoy en día aceptan que sus ordenadores de vez en cuando se estropeen (Ceruzzi 1998, 177-206).

Para las aplicaciones aeroespaciales hubo que abandonar este modelo, ya que un fallo informático significaba que la nave podía estrellarse. La campaña de las fuerzas aéreas llamada «Alta fiabilidad», lanzada a finales de la década de 1950, consiguió este objetivo. Los fabricantes desarrollaron técnicas de control de calidad y cada paso de la fabricación se documentaba rigurosamente. Los aparatos se ensamblaban en «salas estériles» (inventadas en un laboratorio de armamento de Nuevo México), que eran más asépticas que una sala de quirófano. En ellas los trabajadores llevaban trajes que impedían que trozos de piel o de cabello contaminaran las piezas del montaje y había filtros para detectar la más mínima partícula de polvo. También en esta década los químicos desarrollaron formas de producir silicio ultra puro, en el cual se introducían cantidades mínimas y muy precisas de otros elementos con el fin de obtener un material con las propiedades electrónicas deseadas (un proceso llamado «dopaje»). Gran parte de esta actividad estaba localizada en lo que en otro tiempo fuera un valle agrícola al sur de San Francisco y que pronto fue bautizado Silicon Valley por un periodista local. La compañía Fairchild Semiconductor, donde trabajaba Robert Noyce, era el centro de aquella actividad creativa. Allí, además de desarrollar las técnicas de manipulado del silicio

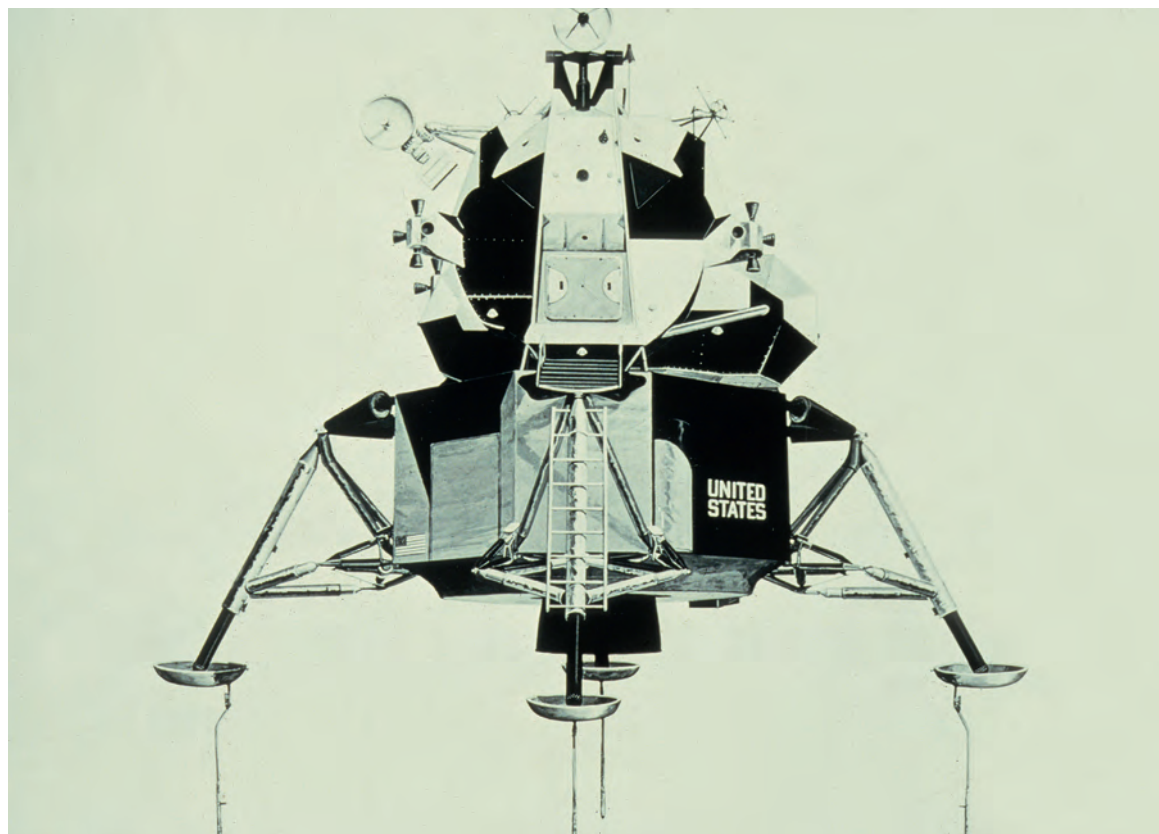
antes mencionadas, los ingenieros concibieron un método de fabricar transistores por medio de fotolitografía. Todos estos avances fueron anteriores a la invención del circuito integrado, pero sin ellos lo que siguió no habría sido posible.

El circuito integrado significaba que en un trozo de material se colocaba más de un mecanismo. Al principio el número de circuitos en un chip era pequeño, unos cinco o seis. Pero ese número empezó a duplicarse, primero una vez al año, después cada 18 meses aproximadamente. Este ritmo se ha mantenido desde entonces y recibe el nombre de Ley de Moore, por Gordon Moore, un colega de Robert Noyce en Fairchild responsable de sentar gran parte de los cimientos materiales que hicieron posibles los avances en la fabricación de chips (Moore 1965). Aquella ley (en realidad una observación empírica) ha impulsado desde entonces la informática y también su relación simbiótica con la industria aeroespacial. En este contexto no es de sorprender que el primer contrato para la fabricación de chips en grandes cantidades fuera con el programa de misiles balísticos Minuteman, de las fuerzas aéreas estadounidenses, para un modelo de dicho misil que voló por primera vez en 1964. Casi inmediatamente después del contrato con Minuteman llegó otro para el ordenador que guiaría a los astronautas del *Apollo* en su viaje a la Luna, en una serie de misiones tripuladas que empezaron en 1968 (Ceruzzi 1998, 182). Para entonces, la Ley de Moore empezaba a tener un impacto significativo en la ingeniería aeroespacial y en otros campos. La última de las misiones *Apollo*, una cita en la órbita terrestre con la cápsula soviética *Soyuz*, fue en 1975. A bordo iba una calculadora de bolsillo fabricada en la compañía de Silicon Valley Hewlett-Packard. Aquella calculadora tenía más capacidad informática que el ordenador encargado de la navegación, diseñado una década antes, cuando el chip era una novedad. Los ejemplos de situaciones similares son numerosos.

Los espectaculares avances hasta ahora mencionados y otros, como las misiones robóticas al espacio, son en gran medida el resultado de la influencia de la Ley de Moore en el diseño aeroespacial, especialmente en naves no tripuladas por humanos (quienes, para mejor o peor, siguen teniendo las mismas necesidades de espacio, comida, bebida y oxígeno hoy que en 1959, cuando se inventó el chip de silicio). La comparación directa de ARPANET con el proyecto *Apollo* deja fuera importantes matices de la historia. Una de sus ironías es que los avances en exploración espacial han influido también en el diseño de las naves. El módulo lunar *Apollo*—aquella nave desgarrada que transportó a dos astronautas por los 100 kilómetros que separan la órbita lunar de la superficie de la Luna— tenía que estar controlado por ordenador, ya que ningún ser humano es capaz de realizar un alunizaje en ausencia de atmósfera, y los controladores de la base de Houston estaban demasiado lejos para resultar de ayuda (Mindell 2008). En las etapas finales del programa *Apollo* los ordenadores de navegación se trasladaron de las naves espaciales a una



Ordenador de guiado Apollo, 1969. Este ordenador desempeñó funciones clave de guiado, navegación y control en las misiones que llevaron a un total de 132 astronautas a la Luna y de vuelta a la Tierra entre 1969 y 1972. Estos ordenadores, junto con los de guiado del Minuteman II, fueron de los primeros del mundo en incorporar circuitos integrados. Junto al ordenador hay un teclado con teclas lo suficientemente grandes para que un astronauta pueda pulsarlas con el traje espacial puesto. Smithsonian Institution.



Módulo lunar. El módulo lunar operaba por completo en el vacío espacial y aterrizó en la Luna mediante sus motores cohete. No admitía tripulación humana y, debido a la gran distancia de la Tierra, tenía que ser controlado por un ordenador a bordo. NASA.

nave experimental de la NASA, para comprobar si ésta se beneficiaba también de esta tecnología. No fue una coincidencia que la NASA escogiera como director del programa al mismísimo Neil Armstrong, el primer hombre que caminó sobre la Luna en 1969, y por tanto el primero cuya vida dependió en última instancia del correcto funcionamiento de un ordenador digital (Tomayko 2000).

Los ensayos de la NASA fueron un éxito, pero las compañías aéreas estadounidenses tardaron en adoptar las nuevas tecnologías. El consorcio europeo Airbus, sin embargo, sí lo hizo, empezando a finales de la década de 1980 con el Airbus A-320. A diferencia del módulo lunar, el avión no requiere comandos eléctricos de vuelo (*fly-by-wire*), pero el uso de un ordenador daba al A-320 una mayor comodidad a la hora de volar y le permitía ahorrar más combustible que sus competidores estadounidenses Boeing y McDonnell-Douglas. Los comandos eléctricos de vuelo, unidos a las modernas cabinas de vidrio o *glass cockpits* (pantallas de ordenador con información de los sistemas del avión) forman hoy parte de todos los aviones comerciales y militares. La Lanzadera Espacial también emplea comandos eléctricos de vuelo en su diseño, ya que sin estos sistemas de control sería poco práctico que la pilotara un humano y esperar que realizara un aterrizaje de precisión después de entrar en la atmósfera a una velocidad de 27.000 kilómetros por hora.

Otra influencia directa de las fuerzas aéreas y de la NASA en la industria informática fue el desarrollo del CAD (siglas de Computer Aided Design, diseño asistido por ordenador). Las fuerzas aéreas patrocinaron un proyecto del Massachusetts Institute of Technology (MIT) que desembocó en el control de las máquinas por una secuencia digital de controles informáticos, codificados en forma de agujeros perforados en una banda de cinta plástica. Los resultados de este trabajo transformaron el funcionamiento de los aparatos no sólo en la industria aeroespacial sino en la metalurgia en general. Al mismo tiempo, ingenieros de varios centros de la NASA llevaban un tiempo empleando ordenadores para que asistieran en los análisis de impacto de cohetes y naves espaciales. Los vehículos lanzadera debían ser lo suficientemente resistentes para almacenar el oxígeno y el combustible así como para soportar las estructuras superiores, al tiempo que resistían la vibración y el impacto del lanzamiento. Además, tenían que ser ligeros. Los ingenieros aeronáuticos llevaban décadas enfrentándose a este problema; en una compañía aérea, por cada especialista en aerodinámica en nómina podía muy bien haber diez ingenieros trabajando en análisis de impacto. Su misión era asegurar que la nave era lo suficientemente resistente como para sobrevivir a un vuelo y al mismo tiempo lo bastante ligera como para despegar del suelo. La NASA fue la primera en poner

en marcha la investigación informática en este campo y entre los resultados obtenidos estuvo un programa de análisis de impacto generalizado llamado NASTRAN: abreviatura de: NASA Structural Analysis (análisis estructural de la NASA), basado en el entonces ya popular lenguaje de programación FORTRAN. Desde entonces es de uso común en la industria aeroespacial.

La adopción por parte de los aviones comerciales del sistema de comandos eléctricos de vuelo instaurado por el programa Apollo puso en primer plano la cuestión de la fiabilidad. La invención del chip de silicio, en combinación con las iniciativas «Alta Fiabilidad» de las fuerzas aéreas, contribuyó en gran medida a hacer de los ordenadores una herramienta apropiada para la navegación espacial, pero la fiabilidad seguía estando en tela de juicio. Si los ordenadores del *Apollo* fallaban durante el vuelo los astronautas podría regresar a la base guiados por un ejército de controladores en Houston. Ningún ordenador del programa Apollo falló nunca, pero durante el vuelo del *Apollo 13* en 1970 la nave se quedó prácticamente sin energía eléctrica y la tripulación no habría sobrevivido de no ser por los controladores de Houston, que les aconsejaron seguir adelante con el aterrizaje. Tener todo un equipo de controladores en tierra pendientes de cada vuelo comercial es obviamente inviable, por eso la Lanzadera Espacial, cuya misión es proporcionar acceso al espacio de manera rutinaria, fue diseñada de un modo diferente. Para el A-320 Airbus desarrolló un sistema de tres ordenadores eléctricos que «votan» antes de cada acción. El fallo durante el vuelo de uno de ellos sería contrarrestado por los votos de los otros dos, de manera que la nave siempre podría aterrizar sana y salva. La Lanzadera tiene cinco ordenadores, de manera que el fallo de uno de ellos no impediría que se completara la misión. El quinto ordenador está para el supuesto de que se produzca un error de *software* (Tomayko 1987, 85-133). Esta medida de seguridad se ha generalizado en el diseño aeronáutico. Muchas

naves espaciales también la adoptan, pero en menor medida, especialmente si la tripulación no es humana.

Si la incorporación de herramientas informáticas a la aeronáutica comercial ha transformado los aviones de pasajeros, la situación en tierra no ha progresado más allá del tubo de vacío. El tráfico aéreo comercial es hoy muy seguro y dicha seguridad depende de que los controladores de vuelo dirijan el tráfico por autopistas aéreas virtuales. Puesto que Estados Unidos fue pionero en esta actividad, invirtió grandes cantidades de dinero en una tecnología que se basa parcialmente en ordenadores terrestres relativamente anticuados, con comunicaciones con los pilotos vía radio VHF en la frecuencia AM, también una tecnología anticuada. La llegada del GPS (Global Positioning System, sistema de posicionamiento global) —un ejemplo tan válido como cualquier otro de la Ley de Moore— debería permitir a los controladores de vuelo prescindir de gran parte de su infraestructura y sustituirla con información a bordo enviada directamente a los pilotos desde satélites. En otras palabras, en lugar de que los controladores estén pendientes de la localización y la ruta de un avión, los mismos pilotos lo harán, con un método que no afecta la seguridad del vuelo y sí aumenta la capacidad de las vías aéreas. Los pilotos obtendrían información acerca de su posición y del tráfico que pueda interferir en su ruta empleando ordenadores a bordo que procesarían información procedente de la constelación GPS o de otros satélites de navegación, además de otros satélites y de unas cuantas estaciones de tierra. Esto está comenzando a suceder, pero es posible que Estados Unidos sea el último en implantar este sistema.

Si existe un denominador común a todas estas historias sería el de cómo sacar mejor partido de las destrezas humanas frente a las de los ordenadores, ya sea en tierra, en el aire o en el espacio. Se trata de una cuestión aún por zanjar, puesto que está en relación directa con la gradual sofisticación y miniaturización de los ordenadores, lo que obviamente implica que el aparato por sí solo puede desarrollar funciones que antes correspondían a humanos. Pero no es tan sencillo. Los ordenadores que operan en tierra también están mejorando. Los seres humanos tienen hoy las mismas limitaciones físicas que los tripulantes del programa Apollo, pero poseen un conocimiento mucho más profundo de la naturaleza de la navegación espacial y de sus necesidades.



Airbus A-320. Los controles por ordenador desarrollados para hacer aterrizar al módulo lunar se transfirieron al avión comercial. El Airbus A-320, que empezó a operar a mediados de la década de 1980, fue el primer modelo en adoptar esta tecnología fly-by-wire. Hoy todos los aviones comerciales la llevan. Lufthansa.

Necesidades de la informática aeroespacial

Llegados a este punto merece la pena retroceder unas cuantas décadas y examinar algunos aspectos específicos de la navegación espacial y cómo la «informática», en el sentido más amplio de la palabra, está relacionada con ella.

La patente de los hermanos Wright para su avión de 1903 era para un sistema de control de vuelo y no para elevación, sustentación o propulsión. Con la nave espacial ocurre lo mismo; en ellas y en los misiles guiados el control es tan importante como la propulsión. Los misiles guiados se controlan como los aviones, aunque no llevan piloto huma-



Consola de Control de Tráfico Aéreo, en uso hasta la década de 1990. El sistema de Control de Tráfico Aéreo de Estados Unidos compartía muchas características con el sistema de defensa aérea SAGE. Debido a la gran inversión inicial que hizo Estados Unidos en estos sistemas, permanecieron en uso aun cuando estaban ya técnicamente obsoletos. Esta pantalla circular de radar, derivada de un sistema de defensa de la Segunda Guerra Mundial, se empleó en control de tráfico aéreo comercial en Estados Unidos hasta entrada la década de 1990. Smithsonian Institution.

no. Las naves espaciales se mueven en un entorno diferente y por tanto sus necesidades de control también lo son. Un avión o un misil guiado deben tener los motores funcionando constantemente para vencer la resistencia de la atmósfera, mientras que el movimiento hacia delante de las alas en el aire genera una fuerza de sustentación que contrarresta la de la gravedad. Un cohete, en cambio, contrarresta la fuerza de la gravedad no por sustentación sino por empuje. Y una vez sale al espacio exterior la resistencia atmosférica desaparece. Llegado este punto los motores de la nave se apagan. Así, en muchas misiones espaciales, los motores del cohete sólo están en funcionamiento durante una pequeña fracción del tiempo total de navegación. Una nave espacial sí requiere sistema de control, en cambio, pero de manera diferente dependiendo de la fase de la misión en que se encuentre. Durante la fase inicial de vuelo con motor, que puede durar sólo unos pocos minutos, la clave está en alinear el vector de empuje de cohete con el centro de gravedad del vehículo lanzadera. La configuración de la mayoría de los cohetes, con los motores abajo y los tanques de combustible y la carga útil arriba, es inestable. El vehículo tiene a «querer» volcarse y, por un instante, podrá hacerlo si el empuje no está perfectamente guiado durante el ascenso. Una vez lograda la estabilidad en esta fase inicial, el sistema de guiado del vehículo puede dirigir el empuje de manera que se desvíe de esta alineación, primero ligeramente y luego de forma cada vez más marcada conforme vaya ganando velocidad. Esto hará que la nave se ladee hasta alcanzar un ángulo óptimo en el que el empuje no sólo contrarresta la gravedad sino que también propulsa el cohete en sentido horizontal: para entrar en órbita, regresar a la Tierra o abandonar por completo la atmósfera terrestre.

Controlar el empuje de un cohete en esta primera fase de una misión espacial recibe el nombre de «guiado», aunque parece no haber consenso sobre este término en el mundo de la aeronáutica. Además, este sistema de guiado es necesario durante la práctica totalidad de la trayectoria de un misil de combustión, en la cual los motores están operativos casi todo el tiempo.

Una vez que la nave espacial alcanza la velocidad deseada, puede dirigirse a su destino con una trayectoria «balística», llamada así por su parecido a una roca lanzada al aire. Ello en el supuesto de que la velocidad alcanzada sea la correcta en el momento en que se apagan los motores. En caso contrario, o bien los motores principales o bien otros auxiliares se usan para corregir la trayectoria. Esta operación recibe el nombre de «navegación», aunque también aquí la definición no está consensuada. De nuevo a diferencia de los barcos o de los aviones, en misiones de larga distancia una nave espacial puede disparar sus cohetes sólo ocasionalmente y no de forma continua (a excepción de los sistemas de propulsión por iones y eléctrico). Pero el proceso es el mismo: determinar si el rumbo es correcto y, de no ser así, arrancar los motores para modificar la velocidad según sea necesario.

Por último, una nave que surca el espacio no está sujeta a las fuerzas de la atmósfera. Una vez los motores se han detenido es libre de emprender cualquier dirección y mantendrá el mismo rumbo independientemente de adónde apunte. En la práctica una misión requiere que la nave se oriente en una dirección específica: que coloque sus paneles solares en dirección al Sol, que apunte con su cámara a un punto en la Tierra, a una antena, etc. El proceso de orientar una nave con sus ejes x , y y z en el espacio se llama «control». Una nave espacial adquiere control empleando cohetes de motor de pequeño empuje, bobinas magnéticas, ruedas impulsoras, gradientes de gravedad y otras herramientas aún más exóticas. El término «control» también incluye aspectos operativos de la misión espacial, tales como la puesta en funcionamiento de una cámara, la activación de instrumentación, preparar un vehículo para su captura por la atmósfera de otro planeta, etc. Estas acciones pueden hacerse de forma automática, por los miembros de la tripulación a bordo o por estaciones de «control de misión» emplazadas en tierra firme.

La aeronave de los hermanos Wright tenía un diseño inestable y requería atención constante por parte del piloto. Trasladar el estabilizador horizontal a la parte trasera del avión proporcionaba una mayor estabilidad, de la misma manera que las plumas de la cola estabilizan una flecha. Pero el control del vuelo aeronáutico aún era una tarea complicada. Con objeto de ayudar al piloto a mantener el control de aparato, el inventor estadounidense de principios del siglo xx Elmer Sperry desarrolló un sistema de giróscopos que aumentaba la estabilidad del aparato y facilitaba el pilotaje del mismo. Esta combinación de colocación de las superficies de control aerodinámico en la popa con un

sistema autoestabilizador basado en giróscopos se trasladó al diseño y fabricación de cohetes. Los sistemas de guiado de los cohetes modernos todavía incorporan una versión, más compleja, eso sí, del invento de Sperry. De estas mejoras, una resultó de especial importancia para el sistema de guiado de cohetes y procedía del programa alemán V-2: el diseño de un giróscopo pendular para medir el tiempo integral de aceleración, que (según los cálculos de Newton), indica la velocidad de la nave (Mackenzie 2000).

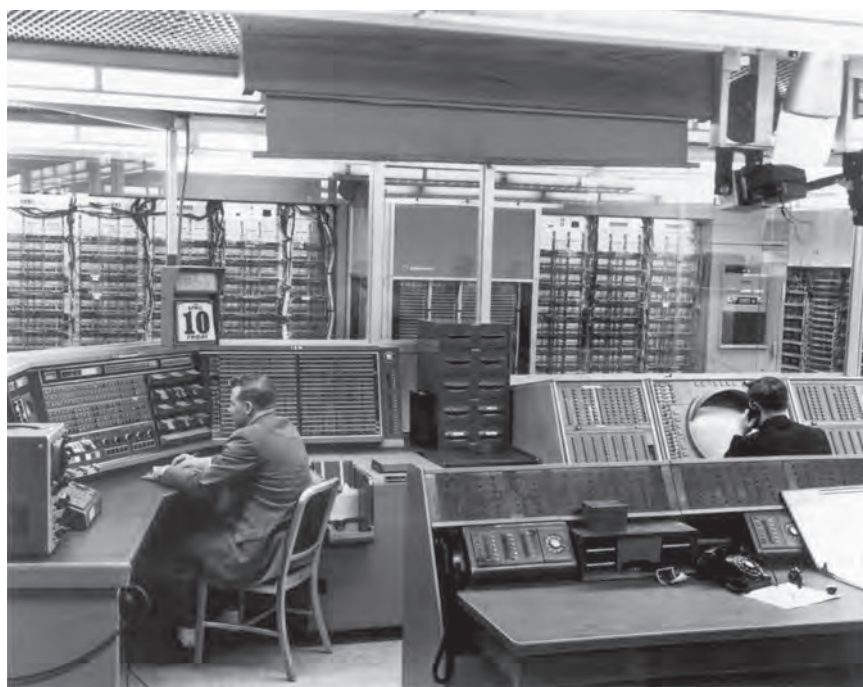
Durante la fase de vuelo con motor el guiado debe realizarse a velocidades proporcionales a la acción del cohete. Esto deja fuera cualquier intervención de humanos situados en el punto de lanzamiento, aparte de decisiones tales como destruir un cohete que se ha salido de rumbo. Las funciones de control también pueden desempeñarlas sistemas instalados a bordo, pero si no existe urgencia por orientar una nave puede hacerse por medio de comandos dictados desde tierra. La navegación a menudo puede ser lenta, con tiempo suficiente para procesar datos recogidos por radar a través de potentes ordenadores centrales, que a continuación pueden transmitir las instrucciones según éstas sean necesarias. Así, mientras que el guiado se realiza fundamentalmente a bordo de la nave por medio de giróscopos y acelerómetros que operan sin comunicación con el exterior, la navegación y el control pueden combinar información procedente de los sistemas a bordo con señales de radio a y desde las bases situadas en tierra. Algunos de los primeros misiles balísticos se guiaban también por radio desde tierra, aunque a velocidades a tiempo real y sin intervención

humana directa en el momento de lanzamiento. Esta modalidad de guiado por haz direccional ha caído hoy en desuso.

Trasladar las señales procedentes de un giróscopo integrado o un acelerómetro requería de lo que hoy llamamos «computación» o «cómputo». Los primeros sistemas empleaban sistemas electromecánicos de marchas y repetidores. Se trataba de ordenadores analógicos con un diseño que era «análogo» a las condiciones de vuelo que debía controlar. El V-2, por ejemplo, utilizaba un giróscopo pendular para computar la integral de aceleración, informando así de la velocidad; a determinada velocidad el motor se apagaba para alcanzar un objetivo predeterminado. Estos primeros mecanismos fueron más tarde sustituidos por sistemas electrónicos que empleaban tubos de vacío. Sin embargo los tubos de vacío, aunque de acción rápida, seguían siendo intrínsecamente frágiles e inestables, y sólo se utilizaban en contadas ocasiones.

Los sistemas electrónicos mejoraron sensiblemente con la llegada de las tecnologías de estado sólido, primero el transistor y luego el circuito integrado, ya descritos. Estos circuitos no sólo eran pequeños y resistentes, también permitían diseñar sistemas de control digitales en vez de analógicos y así aprovechar la mayor flexibilidad del ordenador digital. La tecnología digital se emplea ya extensamente no sólo en la industria aeroespacial, también en los nuevos misiles guiados así como en los aviones comerciales y militares. Aunque acertadamente calificado de «revolución», el cambio tardó tiempo en producirse y los primeros ordenadores digitales no aparecieron hasta mediados de la década de 1960, con sistemas como el proyecto Géminis.

Mucho antes de eso, sin embargo, el ordenador digital tuvo una gran influencia en los controles de la aviación terrestres. El V-2 operaba a excesiva velocidad para ser controlado —o rastreado o interceptado— por un ser humano durante el vuelo. Los nuevos aviones no eran tan rápidos, pero igualmente dificultaban la capacidad de los humanos de controlarlos. Hacia 1950 se empezó a aceptar que la computadora electrónica digital, situada en tierra, allí donde su peso y su tamaño no revestían tanta importancia, podía resolver el problema. El proyecto Whirlwind, desarrollado por el Massachusetts Institute of Technology, consiguió dirigir con éxito un avión de las fuerzas aéreas para que interceptara otro en Cape Cod en abril de 1951. A Whirlwind le siguió SAGE, acrónimo de Semi-Automatic-Ground-Environment (entorno terrestre semiautomático), un gigantesco sistema de radares, ordenadores y redes de comunicación que mantenía a Estados Unidos informado acerca de cualquier incursión de la aviación soviética en el Polo Norte. Voces críticas argumentan que el SAGE ya estaba obsoleto para cuando se completó, ya que el misil balístico había sustituido al bombardero como método de lanzamiento de armas. SAGE no podía proporcionar defensa contra misiles balísticos, pero sí fue una importante inspiración para muchos sistemas de control terrestre posteriores, incluidos los empleados hoy día por



Ordenador para la defensa aérea SAGE, c. 1959. El sistema de defensa aérea SAGE influyó enormemente el curso de la industria informática, aunque cuando se hizo operativo, a finales de la década de 1950, muchos lo consideraran ya obsoleto. Fue el primero en emplear visualización de gráficos, en operar a tiempo real y en permitir el funcionamiento interactivo y en red. Smithsonian Institution.



Sistema de reservas manual de American Airlines, c. 1957. Uno de los beneficios derivados de la tecnología SAGE fue el sistema de reservas aéreas. American Airlines e IBM desarrollaron el sistema "SABRE", todavía en uso, para reemplazar los trabajosos métodos manuales ilustrados en la fotografía. Smithsonian Institution.

la Administración Federal de Aviación de Estados Unidos para gestionar el tráfico aéreo comercial (Ceruzzi 1989).

Llegada la década de 1960 las operaciones espaciales ya estaban controladas en gran medida desde tierra. En el diseño inicial del proyecto Mercury, por ejemplo, el papel del astronauta era prácticamente el de simple pasajero, con estaciones de control terrestres repartidas por todo el mundo encargadas de guiar la misión, hasta el punto de que las primeras cápsulas espaciales del proyecto ni siquiera tenían ventanas. A partir de entonces las naves espaciales con tripulación humana empezaron a ganar en autonomía, pero ninguna de ellas tiene autorización para maniobrar sin información directa de los controladores en tierra. El rescate de la tripulación del *Apollo 13* en 1970 puso de manifiesto la suma importancia de dichos controles. Hoy, la mayoría de las operaciones espaciales, desde la Estación Lanzadera y Espacial hasta los satélites de comunicación comerciales y las misiones militares y de investigación científica sin tripulación, requieren mayores infraestructuras terrestres que la aviación comercial y militar.

SAGE fue diseñado para detectar aviones enemigos. Una década más tarde Estados Unidos empezó a desarrollar el BMEWS (Ballistic Missile Early Warning System), un sistema de alerta temprana de misiles balísticos. La defensa aérea del continente se consolidó con la creación de una infraestructura llamada NORAD (North American Aerospace Defense Command, Comando de defensa aeroespacial de Norteamérica) en Colorado Springs, Colorado, donde ordenadores y humanos vigilan de forma continuada el espacio aéreo. La defensa contra misiles balísticos continúa

siendo un objetivo escurridizo. En el presente los esfuerzos están agrupados bajo el término Defensa Nacional contra Misiles, que ha desarrollado algunos prototipos de *hardware*. Se han instalado unos cuantos sistemas diseñados para interceptar misiles de corto alcance en varios lugares del planeta. Los ordenadores desempeñan un papel crucial en dichos esfuerzos: ayudan a detectar el lanzamiento de un misil, a rastrear su trayectoria, a discernir objetivos legítimos de meros señuelos y a dirigir los interceptores. Estas actividades requieren de una potencia informática gigantesca, así como de elevadas capacidades de procesamiento. La defensa contra misiles está impulsando el campo de la informática hasta extremos desconocidos para el usuario común, a años luz de los teléfonos móviles, ordenadores portátiles o reproductores multimedia.

Para la detección de señales de inteligencia procedentes de satélites también se han desarrollado sistemas de control igualmente elaborados y costosos. Aunque los detalles acerca de los mismos son información clasificada, podemos afirmar que gran número de los sistemas militares estadounidenses son controlados desde infraestructuras terrestres localizadas en las proximidades de Colorado Springs, y los relacionados con la navegación espacial, en Houston. Todos ellos pueden considerarse descendientes legítimos del proyecto Whirlwind.

Por último cabe añadir una observación relativa a la naturaleza del control terrestre frente al control a bordo. SAGE son las siglas de control terrestre semiautomático. El prefijo «semi» se incluyó para dejar bien claro que seres humanos participaban activamente del proceso y que ningún sistema informático podría actuar por su cuenta, sin intervención humana. Del mismo modo, en los programas de navegación espacial, los planes iniciales de poner en órbita naves controladas por completo desde tierra no llegaron a ponerse en práctica. Los primeros diseños del proyecto Mercury fueron modificados, al principio accediendo a presiones por parte de los astronautas, más tarde una vez los primeros vuelos pusieron de manifiesto que era absurdo que el tripulante tuviera un papel del todo pasivo durante el vuelo. El deseo de la intervención humana está también presente en la gestión de la Lanzadera Espacial, que no puede operar sin un piloto de carne y hueso.

El futuro

De lo discutido hasta ahora cabe deducir que no es posible establecer una comparación simple entre los avances en informática y en navegación espacial desde 1958. No obstante, a los productores del programa de televisión «Nerds, 2.0.1.» no les faltaba razón. Internet ha experimentado una rápida difusión entre la sociedad que la navegación espacial no ha sido capaz de igualar. Un factor no mencionado en el programa pero que puede ser relevante es una observación hecha por el pionero en redes informáticas Robert Metcalfe. Según éste, el valor de una red de comunicaciones aumenta conforme lo hace el cuadrado del número

ro de sus usuarios (una afirmación bautizada por él mismo como Ley de Metcalfe, en oposición a la Ley de Moore). Así Internet, que aumenta su cifra de usuarios cada día, incrementa su valor mucho más rápidamente que el coste que supone hacer cada una de estas nuevas conexiones. La exploración espacial carece de una ley equivalente, aunque si las sondas espaciales terminan por hallar indicios de vida en otros planetas habrá que reescribir la ecuación.

Una faceta de la historia que suele olvidarse cuando se escribe sobre Internet es que la comunidad aeroespacial estuvo entre las pioneras en la utilización de redes informáticas, pero con objetivos distintos. El sistema SAGE fue la primera red informática a gran escala del mundo, por ejemplo. Y la primera vez que se utilizó una red informática para uso privado, es decir, ni militar ni gubernamental, fue el sistema de reservas aéreas SABRE, desarrollado por IBM para American Airlines a principios de la década de 1960. De hecho, ARPANET se desarrolló en parte para paliar las deficiencias de SAGE. En este último toda la red se volvería inoperativa si se destruía uno de los nodos de control centrales. Con Internet eso no puede ocurrir ya que, por diseño, carece de un punto de control central. La capacidad de Internet de vincular sistemas informáticos distintos mediante un conjunto de protocolos comunes la diferencia de otras redes informáticas aeroespaciales, que a menudo no pueden comunicarse entre sí. Un ejemplo especialmente embarazoso de ello ocurrió recientemente durante el desarrollo por parte de Airbus de su superjumbo, el Airbus-380, el cual empleaba un programa informático CAD llamado CATIA, desarrollado por la compañía francesa Dassault Systèmes. CATIA permitía a ingenieros de distintos laboratorios y plantas industriales trabajar conjuntamente en una serie de «dibujos» o planos virtuales, como si estuvieran en el mismo edificio. Para el A-380, un grupo de diseñadores empleaba una versión modificada de CATIA, y cuando se reunieron las distintas partes para el ensamblaje final en la planta industrial de Airbus en Toulouse, resultó que no encajaban. Boeing también ha experimentado problemas similares a la hora de integrar ensamblajes procedentes de distintos puntos de fabricación para su nuevo avión, el 787 Dreamliner. Aunque para hacer justicia a Airbus y Boeing hay que decir que Internet, tal y como está actualmente configurada, no sería capaz de gestionar las complejidades que implica el diseño de un aeroplano moderno, a pesar de que sí es más operativa que un gran número de nodos repartidos por todo el mundo.

¿Significó el proyecto Apollo de la NASA, por muy impresionante que fuera desde el punto de vista de la ingeniería, un punto muerto en la carrera espacial? ¿Y

es la red de comunicaciones desarrollada por la agencia gemela de la NASA, ARPA, la tecnología definitiva de la era moderna? Ninguna de las dos preguntas es sencilla de responder. Ambas tecnologías han crecido en una relación simbiótica y seguirán haciéndolo en el futuro. El concepto de ordenador como un agente dotado de inteligencia artificial al servicio de la humanidad ha dado paso al de la computadora como un artefacto diseñado para «complementar el intelecto humano», en palabras del pionero de la informática Douglas Engelbart. Engelbart es especialmente conocido por la invención del ratón informático, pero también por ser uno de los primeros científicos en identificar el lugar que ocuparían los ordenadores en nuestras vidas. Antes de inventar el ratón Engelbart trabajó en el centro de investigación de la NASA, Ames, en Mountain View, California y más tarde para una firma de redes informáticas propiedad de la compañía aeroespacial McDonnell-Douglas. Así que conocía bien las limitaciones del mundo real y también el verdadero potencial de la informática en red y de sus aplicaciones en la industria aeroespacial.

Las limitaciones del cuerpo humano continuarán siendo un lastre en el progreso de la exploración del espacio. Con el conocimiento que tenemos hoy día de las leyes físicas resulta difícil concebir expediciones humanas más allá de la órbita de Marte, incluso con las extrapolaciones más optimistas de los recientes cohetes con motores de propulsión química. Si continúa la tendencia actual, hacia el año 2030 los ordenadores incorporarán tantos circuitos como neuronas hay en el cerebro humano. Si tenemos en cuenta esta equivalencia, entonces podría concebirse transferir la naturaleza de la conciencia humana a un ordenador, que a su vez podría explorar el Cosmos libre de las limitaciones que trae consigo la presencia humana, la cual, de momento, todavía es necesaria. Tal es el argumento del inventor Ray Kurzweil, quien está convencido de que dicha transferencia de conciencia es inevitable (Kurzweil 1999). Claro está que ello pasa por que esta transferencia se produzca en realidad. Ya hemos visto que las predicciones iniciales sobre inteligencia artificial resultaron ser en exceso optimistas. Disponer de más y más circuitos puede no ser suficiente para que un ordenador cruce el umbral que separa la «inteligencia», sea cual sea su definición, de la «conciencia». Es aconsejable dejar estas especulaciones a los escritores de ciencia-ficción. Podemos sentirnos decepcionados por el hecho de que la exploración del espacio presente, a día de hoy, tantas limitaciones, pero es difícil conservar esa impresión mucho tiempo a la vista de todos los otros emocionantes avances en aeronáutica espacial que nos sorprenden cada día.

Bibliografía

- Abbate, Janet. *Inventing the Internet*. Cambridge, Massachusetts: MIT Press, 1999.
- Ceruzzi, Paul. *Beyond the Limits: Flight Enters the Computer Age*. Cambridge, Massachusetts: MIT Press, 1989.
- , *A History of Modern Computing*. Cambridge, Massachusetts: MIT Press, 1998.
- Dick, Steven J., y Roger Launius, eds. *Societal Impact of Spaceflight*. Washington, DC: NASA, 2007.
- Kranz, Gene. *Failure is not an Option*. Nueva York: Berkeley Books, 2000.
- Kurzweil, Ray. *The Age of Spiritual machines: When Computers Exceed Human Intelligence*. Nueva York: Viking, 1999.
- Mackenzie, Donald. *Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance*. Cambridge, Massachusetts: MIT Press, 1990.
- McCorduck, Pamela. *Machines Who Think: A personal Inquiry into the History and Prospects of Artificial Intelligence*. San Francisco, California: W.H. Freeman, 1979.
- Mindell, David. *Digital Apollo: Human and Machine in Spaceflight*. Cambridge, Massachusetts: MIT Press, 2000.
- Moore, Gordon. "Cramming More Components onto Integrated Circuits." *Electronics*, 19 de abril de 1965, 114-117.
- Noble, David F. *Forces of Production: A Social history of Industrial Automation*. Nueva York: Oxford University Press, 1986.
- Norberg, Arthur, y Judy O'Neill. *Transforming Computer Technology: Information Processing for the Pentagon, 1962-1986*. Baltimore, Maryland: Johns Hopkins University Press, 1996.
- Segaller, Stephen. *Nerds: A Brief History of the Internet*. Nueva York: TV Books, 1998.
- Tomakyo, James E. *Computers in Spaceflight: the NASA Experience*. Nueva York: Marcel Dekker, Inc., 1987.
- , *Computers Take Flight: A History of NASA's Pioneering Digital Fly-By-Wire Project*. Washington, DC: NASA, 2000.

internet: su evolución y sus desafíos

JANET ABBATE

Internet, una red global de redes, es un sistema notablemente complejo desde el punto de vista técnico, construido a partir de las contribuciones creativas de científicos de todo el mundo a partir de la década de 1950 y hasta el presente. A lo largo de su evolución, Internet y otras redes han sido impulsadas por gobiernos, investigadores, educadores e individuos en general como herramientas para satisfacer un gran número de necesidades de la población. La combinación de políticas estatales e improvisaciones de base (a escala local) ha dado lugar a beneficios sociales que incluyen el acceso generalizado a ordenadores y a información, una colaboración más estrecha entre científicos, crecimiento económico, formación de *comunidades virtuales* y una mayor capacidad para mantener lazos sociales a larga distancia; también a la democratización de los contenidos creativos y al activismo político y social en línea. El rápido crecimiento de Internet también ha generado crisis técnicas, tales como la congestión y la escasez de dominios, así como dilemas sociales, incluidas actividades ilegales o malintencionadas, y continuas *discriminaciones digitales* basadas en diferencias de ingresos, procedencia, edad, sexo y educación. Estos problemas siguen requiriendo soluciones creativas por parte de científicos, legisladores y ciudadanos.

Del desarrollo técnico de Internet hay varios aspectos a destacar. En primer lugar, desde 1950 hasta el presente se ha producido un aumento continuado en el volumen de las redes de datos y en la variedad de servicios que éstas

ofrecen. El rápido crecimiento y la diversidad han obligado a los diseñadores de redes a superar incompatibilidades entre sistemas y componentes informáticos, a gestionar el tráfico de datos para prevenir la congestión y el caos y a llegar a acuerdos internacionales sobre estándares técnicos. Estos desafíos han cristalizado en logros fundamentales en áreas de investigación tales como sistemas operativos y la teoría de colas (modelo matemático para el estudio de fenómenos de espera). Una segunda tendencia ha sido el diseño de funciones de las redes en forma de *capas de enlace*, cada una de las cuales se comporta según un *protocolo estándar* (una serie de normas para interacción implementada en *software* o en *hardware*). Este diseño por capas reduce la complejidad del sistema de redes y minimiza el grado de estandarización necesario para su funcionamiento, lo que facilita que las redes puedan unirse a Internet. Un tercer e importante rasgo del desarrollo técnico de Internet ha sido un proceso de diseño inusualmente descentralizado y participativo. Ello ha abierto el sistema a innovaciones procedentes de una variedad de fuentes y ha fomentado la colaboración informal desde todas las zonas de planeta. En los siguientes apartados se describen algunos de los hitos principales en la evolución de Internet y sus predecesoras.

Los inicios: primeros terminales en redes

Las primeras computadoras electrónicas digitales, inventadas durante la Segunda Guerra Mundial y comerciali-

zadas inmediatamente después, eran máquinas solitarias; no estaban diseñadas para interactuar con sus usuarios ni con otras computadoras. Al cabo de unos pocos años, sin embargo, los ingenieros informáticos comenzaron a experimentar con maneras de acceder a distancia a los ordenadores o de transmitir datos de una máquina a otra. Las redes de datos de la década de 1950 y principios de la de 1960 eran sistemas concebidos para conectar terminales a ordenadores, antes que ordenadores entre sí. Los experimentos realizados con terminales en red resultaron ser un área de investigación llena de desafíos para los científicos, pero también una respuesta a la realidad política y económica del momento, marcada por la Guerra Fría y por el crecimiento global de la economía, los transportes y las comunicaciones.

La ciencia informática en Estados Unidos fue en gran parte una creación del ejército y reflejaba la rivalidad del país con la URSS. Así por ejemplo, un importante avance estadounidense en la década de 1950 fue el proyecto SAGE (un sistema informatizado de alerta temprana de detección de misiles enemigos). Cada centro SAGE tenía un ordenador IBM que recibía datos por medio de líneas telefónicas conectadas a instalaciones de radar y bases militares. Una de las tecnologías clave desarrolladas por los laboratorios Bell AT&T para SAGE fue el *módem*, que sirve para convertir datos informáticos digitales en señales analógicas que pueden enviarse por la red telefónica. AT&T empezó a comercializar los módems en 1958 y durante varias décadas éstos fueron el principal acceso en red para el usuario común.

La demanda de terminales en red se vio impulsada por otro hito tecnológico de principios de la década de 1960: los sistemas operativos *de tiempo compartido*. Inventados de forma independiente en 1959 por Christopher Strachey en Gran Bretaña y por John McCarthy en Estados Unidos, los sistemas operativos de tiempo compartido permitían que múltiples usuarios operaran con programas distintos en un solo ordenador y de forma simultánea. Debido a que el coste del ordenador podía repartirse entre varios usuarios, el sistema de tiempo compartido permitía a individuos usar un ordenador de forma interactiva durante largos periodos de tiempo, en lugar de estar limitados a operar con un solo programa, y también a recibir los resultados de forma inmediata. Los servicios comerciales de tiempo compartido se aprovecharon de este ahorro para proporcionar acceso informático a precios asequibles a gran cantidad de clientes del mundo académico y de los negocios. A mediados de la década de 1960 los servicios comerciales de tiempo compartido ya estaban desarrollando sus propias redes de datos con objeto de ofrecer a sus clientes acceso a bajo coste a sus ordenadores.

El capitalismo global y el crecimiento de los sistemas de transporte y comunicaciones dieron impulso a la creación de terminales de redes comerciales y de gran escala. A principios de la década de 1960 las industrias que

manejan gran cantidad de datos, tales como la aviación o las compañías financieras, empezaron a desarrollar redes conjuntas que permitieran a distintas empresas compartir información. Por ejemplo, a principios de los sesenta American Airlines e IBM crearon SABRE, un sistema de reservas en línea (basado en el proyecto de IBM para SAGE), el cual conectaba 2.000 terminales de Estados Unidos a un ordenador central. De manera similar, el NASDAQ o Asociación Nacional de Cotización Automatizada de Operadores de Valores de Estados Unidos, creó una red para cotizaciones de acciones en 1970. En un ejemplo temprano de la colaboración internacional en redes informáticas, una cooperativa de líneas aéreas llamada SITA (Société Internationale de Télécommunications Aéronautiques) construyó una red en 1969 empleando la técnica de conmutación de paquetes o *packet switching* (véase más adelante). La red de SITA gestionaba el tráfico aéreo de 175 compañías a través de centros informáticos situados en Ámsterdam, Bruselas, Fráncfort, Hong Kong, Londres, Madrid, Nueva York, París y Roma (SITA 2006). Dichas redes comerciales y financieras contribuyeron a acelerar la globalización de la economía.

Redes de investigación

Las redes de terminales se basaban en un modelo radial (*hub-and-spoke*) relativamente simple que conectaba a numerosos usuarios a un solo ordenador central. Entre finales de la década de 1960 y finales de la de 1970 ingenieros informáticos desarrollaron nuevas redes más complejas que conectaban múltiples ordenadores entre sí. Al experimentar con nuevas tecnologías, los investigadores buscaban romper las barreras que entrañaba compartir datos entre ordenadores con sistemas operativos distintos. Los científicos y sus patrocinadores —los gobiernos— vieron una triple promesa en la creación de redes informáticas: la capacidad de compartir ordenadores escasos y costosos, lo que aumentaría el acceso al tiempo que reducía los gastos; la posibilidad de compartir datos y trabajar en colaboración con colegas de otros lugares y la oportunidad de progresar en la teoría y la práctica del uso de los aparatos.

Tres de las primeras redes de investigación más importantes fueron ARPANET (Estados Unidos, 1969), la NPL Mark I (Gran Bretaña, 1969) y CYCLADES (Francia, 1972). Una de las innovaciones clave de estas redes experimentales fue una técnica de comunicación llamada *packet switching* o conmutación de o por paquetes. Hasta entonces los sistemas de comunicación existentes, como el teléfono o las terminales de redes, tenían circuitos establecidos entre las dos terminales de la conexión. Por el contrario, una red de conmutación por paquetes divide los datos de manera que puedan ser transmitidos en pequeñas unidades llamadas «paquetes» que son enviadas de forma individual, compartiendo los circuitos en red con otras conexiones. La conmutación por paquetes permite utili-

zar los nexos de comunicación de manera más eficaz, con el consiguiente ahorro económico. Además, los paquetes de una misma conexión pueden enviarse a su destino por rutas diferentes, haciendo posible distribuir el tráfico en múltiples vías o reaccionar a una interrupción en el funcionamiento de parte de una red redirigiendo el tráfico a otra. El concepto de conmutación por paquetes fue inventado de forma independiente a principios de la década de 1960 por Paul Baran en Estados Unidos y Donald Davies en Gran Bretaña. Davies puso la técnica en práctica en la red de un nodo Mark I en el National Physical Laboratory. En Estados Unidos, la agencia DARPA (siglas de Defense Advanced Research Projects Agency, Agencia para proyectos de investigación de defensa avanzados) patrocinó la primera gran red de conmutación por paquetes: ARPANET. Uno de los teóricos que participó en este proyecto fue Leonard Kleinrock, responsable del desarrollo de algunos de los primeros métodos de análisis del comportamiento de las redes de conmutación por paquetes. En Francia, Louis Pouzin fue pionero de la red sin conexiones o en la red de conmutación de paquetes con *datagramas* CYCLADES. Las redes de datagramas eran más sencillas que las que funcionaban por conexiones, como ARPANET, y esta sencillez hacía más factible interconectar redes diferentes, lo que supuso un importante paso en el desarrollo de una Internet global. Tal y como lo explica Pouzin: «Cuanto más sofisticada es una red, menos probable es que sea compatible con otra» (Pouzin 1975). Experimentos realizados con *internetworking* (conexión de múltiples redes) ya se estaban llevando a cabo a principios de la década de 1970. Por ejemplo, la red NPL se conectó a CYCLADES en 1974, y en 1976 ambas se conectaron a la nueva Red Informática Europea o EIN, desarrollada a partir de un grupo de estudio de ciencia y tecnología de la entonces Comunidad Económica Europea, el cual recomendaba la creación de una red multinacional que ayudara a los países miembros a compartir recursos informáticos y promover la investigación en la ciencia informática. En 1976 EIN proporcionaba servicios de red a diez países, con nodos en Italia, Francia, Suiza y Gran Bretaña (Laws y Hathaway 1978). La convergencia de los sistemas de redes reflejaba así la convergencia entre Estados cooperantes.

Además de la conmutación por paquetes, ARPANET puso en marcha una serie de técnicas experimentales. Dicha red conectaba a investigadores de Estados Unidos que trabajaban en campos tales como tiempo compartido, inteligencia artificial y gráficos. Gracias a lo generoso de las ayudas estatales y a la abundancia de talentos implicados en el proyecto, los diseñadores de ARPANET pudieron experimentar con técnicas prometedoras aunque extraordinariamente complejas. Por ejemplo, en lugar de limitar la red a una sola clase de ordenador, ARPANET incluía variedad de computadoras muy diversas entre sí. Esto impulsó a un equipo de especialistas en informática, estudiantes de posgrado e ingenieros industriales a

buscar nuevas maneras de resolver las incompatibilidades entre ordenadores, y su esfuerzo facilitó en gran medida la creación de la nueva generación de redes informáticas. ARPANET también estaba dotada de una topología *distribuida*, es decir, que constaba de numerosos nodos con múltiples interconexiones, en lugar de un único nodo central. Las comunicaciones distribuidas, descritas por primera vez por Baran en 1964, permitían expandir la carga de información e incrementar potencialmente la fiabilidad, creando múltiples caminos entre dos ordenadores cualesquiera. Sin embargo, la adopción de esta técnica experimental incrementó en gran medida la complejidad del sistema de enrutamiento y forzó a los creadores de ARPANET a analizar y gestionar tipos de comportamiento en red hasta entonces desconocidos. En lo que resultó ser otro paso arriesgado, el diseño en red requería que las operaciones de enrutamiento se descentralizaran y volvieran *adaptativas*: es decir, cada nodo tomaría sus decisiones de enrutamiento de forma independiente y modificaría su comportamiento dependiendo de las condiciones de tráfico de información o de la configuración de la red (por ejemplo, si un nodo adyacente se volvía inoperativo). El diseño descentralizado de ARPANET y su enrutamiento autónomo hacían más difícil analizar el comportamiento en red; al mismo tiempo, estas técnicas contribuirían al futuro éxito de Internet porque permitían que la red creciera sin las limitaciones inherentes a un cuello de botella central. Una de las características más novedosas del proyecto ARPANET no era de tipo técnico, sino organizacional: un proceso de toma de decisiones descentralizado e informal. El desarrollo del *software* para esta red se debió a un equipo de investigadores y estudiantes llamado Network Working Group (Grupo de trabajo para la red). Cualquier miembro del grupo era libre de sugerir una nueva técnica haciendo circular un *boletín de sugerencias*; después de un periodo de debate y experimentación, la sugerencia era modificada, abandonada o adoptada por consenso como un nuevo estándar para la red. Este modelo colaborativo sigue empleándose en Internet (Bradner 1996) y ha ayudado al sistema a crecer y adaptarse, al fomentar el debate de ideas y la participación libre en su desarrollo técnico.

La aplicación de más éxito de las primeras redes experimentales fue con mucho el correo electrónico, cuyo uso se implementó a principios de la década de 1970. La popularidad del mismo fue una sorpresa para los creadores de ARPANET, que habían esperado que las redes dedicadas a la investigación se centraran en aplicaciones más sofisticadas y dedicadas a la computación pura y dura, tales como las matemáticas o el diseño de gráficos. Aunque el *e-mail* se adoptó en parte por su facilidad de uso, su popularidad también ponía de manifiesto la admisión de que la investigación científica dependía de la colaboración humana tanto como del acceso a las máquinas. El correo electrónico resultó ser una oportunidad sin prece-

dentes para la interacción continuada entre colegas geográficamente distantes entre sí.

Aunque no eran accesibles al público en general, las primeras redes de investigación fueron más allá de proporcionar acceso informático a un reducido número de científicos. También aportaron soluciones a los formidables obstáculos técnicos y establecieron herramientas cruciales para innovaciones futuras, incluidas técnicas estandarizadas y una comunidad de investigadores e ingenieros con experiencia en el trabajo en red (Quarterman 1990). Asimismo, los esfuerzos tempranos por construir redes multinacionales y redes de redes sentaron los cimientos para la cooperación global, sin la cual no existiría Internet tal y como lo conocemos hoy.

Expansión creciente:

redes patentadas, públicas y de base

Mediada la década de 1970, la aparición de redes de investigación coincidió con otras tres tendencias: los sistemas de redes patentados que ofrecían los fabricantes de ordenadores; las redes de datos públicos desarrolladas por las empresas de telecomunicaciones o PTT y las redes de base, desarrolladas por individuos y a bajo coste. Compañías como IBM habían proporcionado infraestructuras limitadas de redes informáticas desde la década de 1960, pero una vez que la investigación demostró la viabilidad de la conmutación por paquetes, las empresas informáticas empezaron a ofrecer sus propias tecnologías de *packet switching*. Los sistemas más utilizados incluían el Systems Network Architecture de IBM (SNA), Arquitectura de Red de Sistemas, (lanzado en 1974), el Xerox Network Services (1975) y el Digital Equipment Corporation o DECNET (1975). A diferencia de las redes de investigación, estos sistemas patentados tenían múltiples aplicaciones corporativas. Permitían la diversificación geográfica de las áreas de negocio —ya que las operaciones realizadas en las sucursales tenían acceso a los datos necesarios para operar de forma independiente— y una mayor centralización, porque los datos procedentes de operaciones lejanas podían ser supervisados desde las oficinas centrales. De este modo el uso de redes informáticas reflejaba e intensificaba la tendencia hacia una economía globalizadora que se aceleró a partir de 1980.

Aunque los sistemas patentados proporcionaban un servicio fundamental a organizaciones que trabajaban con muchos ordenadores de un solo fabricante, por lo general no eran compatibles con los de otros fabricantes. Ello podía suponer un problema dentro de una única organización y ciertamente un obstáculo para la creación de una red nacional o internacional. Además, estos sistemas patentados estaban controlados por corporaciones privadas y no se adherían a los estándares técnicos públicamente establecidos. Esto era una importante fuente de preocupación fuera de Estados Unidos, donde se concentraban los grandes fabricantes de ordenadores. Para pro-

porcionar una alternativa, en 1974-1975 los operadores de comunicaciones nacionales de Europa, Canadá y Japón anunciaron planes para construir redes de datos que estuvieran al alcance de cualquier usuario independientemente de la marca de su ordenador.

Las PTT, inspiradas en el modelo del teléfono, ofrecían no sólo acceso universal sino también conexiones internacionales. Al darse cuenta de que ello requeriría llegar a un acuerdo en cuanto al protocolo de red a utilizar, en 1975-1976 el Comité Consultor sobre Telegrafía y Telefonía Internacional del Sindicato Internacional de Telecomunicaciones desarrolló un protocolo estándar de conmutación de paquetes llamado X.25, que proporcionaba una conexión fiable, conocida como *circuito virtual*, entre dos puntos de una red, permitiendo a los usuarios de las distintas terminales acceder a recursos en línea sin necesidad de instalar un complejo *software* específico. Entre los primeros en adoptar este nuevo protocolo figuran la red Datapac de Canadá (1977), Transpac en Francia (1978), DDX de Japón (1979), el PSS de la compañía de correos estatal británica (1980) y la multinacional Euronet (1979). Aunque el X.25 fue más tarde reemplazado por otras tecnologías tales como la transmisión de tramas o *frame relay*, proporcionó la base para el rápido desarrollo de redes públicas en todo el mundo y solucionó el caos que suponía el empleo de estándares incompatibles entre sí. Otro influyente esfuerzo de estandarización ocurrido a finales de la década de 1970 fue el modelo de Interconexión de sistemas abiertos (Open Systems Interconnection) creado por la OSI (Organización Internacional de Estándares). Dicho modelo definía las funciones de siete modelos de proceso, desde conexiones básicas entre *hardware* a aplicaciones de alto nivel e interfaces de usuario. Aunque estos estándares fueron muy debatidos (Abbate 1999), la adopción de un modelo común ayudó a ingenieros informáticos y a fabricantes a dar un paso más en la dirección del desarrollo de sistemas de red totalmente compatibles.

Las redes públicas de datos supusieron el primer acceso *online* para gran parte de la población mundial. También patrocinaron nuevos tipos de contenido y servicios que hacían las redes útiles para los usuarios en general, no sólo técnicos. Por ejemplo, a principios de la década de 1980 France Telecom difundió ampliamente el uso de su red Transpac al ofrecer el innovador servicio Minitel: una terminal gratuita que se daba a los usuarios en lugar de un directorio de teléfonos con acceso también gratuito a una guía de teléfonos *online* y una variedad de servicios de pago. Minitel permaneció en uso durante casi tres décadas y dio servicio a casi la mitad de la población francesa. Al gestionarse los pagos mediante la compañía telefónica, fue uno de los primeros gestores de comercio electrónico, incluyendo reservas de avión y tren, compra al pormenor, gestiones bancarias y de bolsa, servicios de información y tableros de anuncios (McGrath 2004).

El desarrollo de redes públicas de datos puso de manifiesto una conciencia emergente —tanto por parte de usuarios individuales como a nivel estatal— de que el acceso a las comunicaciones informáticas era un bien de interés común, un recurso necesario para el ciudadano del siglo XXI. En el cumplimiento de esta misión las redes públicas de datos se vieron complementadas por la tercera tendencia que se dio en este periodo: las redes de base improvisadas. Estas redes de bajo coste utilizaban *software* existente para intercambiar correos y listas de debate entre una comunidad informal de usuarios. Las más conocidas eran USENET, establecida en 1979 y que empleaba protocolos UNIX, y BITNET, creada en 1981 con protocolos IBM. Dichas redes desempeñaron un papel importante a la hora de comunicar entre sí a personas sin acceso a la infraestructura formal de redes.

Diseñando Internet

¿Cómo fue que estos sistemas tan dispares de comunicación de datos se unieron en la red global que hoy llamamos Internet? Aunque en los años setenta del siglo pasado ya se establecieron algunas conexiones entre redes, las incompatibilidades de diseño limitaban sus servicios al intercambio de correos y noticias. Las tecnologías que permiten que un amplio abanico de servicios en red sea compartido por sistemas distintos surgieron del deseo de conectar ARPANET con dos nuevas redes que ésta había construido y que extendían la conmutación de paquetes a las comunicaciones por radio y satélite. Puesto que estos medios de comunicación no tenían las mismas características técnicas que las líneas telefónicas —las conexiones por radio no eran fiables; los satélites introducían el elemento retardo— las técnicas existentes tales como los protocolos X.25 o el original de ARPANET no servían. Por tanto, a principios de 1970 DARPA puso en marcha un Programa de Internet para desarrollar una solución exhaustiva a estos problemas de comunicación.

Otro avance técnico que ayudó a encauzar la demanda para comunicación entre redes fueron las LAN o *redes locales de área*. Ethernet, la más influyente de todas, la inventó en 1973 Robert Metcalfe a partir de una existente llamada Alohanet, creada a su vez por Norman Abramson, Frank Kuo y Richard Binder (Metcalfe 1996; Abramson 1970). Ethernet y Alohanet fueron pioneras de una técnica denominada *random access* o acceso aleatorio, que permitía a múltiples usuarios compartir un canal de comunicación sin necesidad de emplear complejos sistemas de enrutamiento.¹ La sencillez del diseño de acceso aleatorio acercó las redes de área local a un gran número de usuarios. Ethernet se estandarizó y comercializó a principios de los ochenta y pronto se adoptó en universidades, negocios y otras organizaciones. Otros sistemas populares LAN fueron las redes *token ring*, con topología de anillo, inventadas por investigadores de IBM en Zúrich y comercializadas en 1985. La populari-

dad de las LAN promovió la creación de muchas redes nuevas que podían, en principio, conectarse entre sí; pero, al igual que las redes por conmutación de paquetes, estos sistemas de acceso aleatorio no garantizaban una conexión fiable y por tanto no funcionaban bien con los protocolos existentes. Era necesario desarrollar un nuevo sistema.

El proyecto Internet estuvo dirigido por Vinton Cerf y Robert Kahn con la colaboración de ingenieros informáticos de todo el mundo. Además de investigadores estadounidenses del DARPA, Stanford, la Universidad del Sur de California, la Universidad de Hawai, BBN y el PARC de Xerox, Cerf y Kahn contaron con el asesoramiento de expertos en redes del University College London, los grupos NPL y CYLADES y el International Network Working Group o INWG (Cerf 1990). Este último se había fundado en 1972 e incluía representantes de muchas PTT nacionales que trabajaban en el desarrollo de redes de conmutación por paquetes. Al compartir preocupaciones e ideas, este equipo fue capaz de diseñar un sistema que respondería a las necesidades de usuarios con infraestructuras y necesidades diversas.

La arquitectura de Internet constaba de dos elementos principales. El primero era un conjunto de protocolos llamado TCP/IP (siglas de Transmission Control Protocol, Protocolo de control de transmisiones y de Internet Protocol, Protocolo para Internet, respectivamente) (Cerf y Kahn 1974).² El TCP era un ejemplo de *protocolo anfitrión*, cuya función es establecer y mantener la conexión entre dos ordenadores (*anfitriones*) dentro de una red. La idea detrás de TCP era que el protocolo anfitrión garantizara una conexión fiable entre anfitriones incluso si éstos estaban conectados por una red no fiable, por ejemplo de radiopaquete o el sistema Ethernet. Al reducir los requerimientos de fiabilidad de la red, el uso del protocolo TCP abrió Internet a muchas más redes. Con el fin de asegurar conexiones de confianza, TCP estaba diseñado para verificar la llegada en perfectas condiciones de los paquetes, empleando mensajes de confirmación llamados *asentimientos* (*acknowledgments*); para compensar errores retransmitiendo paquetes perdidos o dañados y para controlar el flujo de datos entre anfitriones limitando el número de paquetes en tránsito. Por el contrario, el Internet Protocol realizaba un conjunto de tareas mucho más sencillas que permitían a los paquetes pasar de una máquina a otra conforme circulaban por la red. IP se convirtió en el lenguaje común de Internet, el único protocolo requerido para cualquier red que quisiera unirse a ella: las redes miembro tenían libertad para escoger entre múltiples protocolos para otras capas del sistema (aunque en la práctica la mayoría terminó por adoptar TCP como protocolo anfitrión). En tanto reflejo de la diversidad de necesidades y preferencias de los expertos que participaron en su diseño, la arquitectura de Internet favorecía la variedad y la autonomía entre sus redes miembro.

¹ La versión mejorada de Metcalfe del sistema de acceso aleatorio recibió el nombre de Carrier Sense Access with Collision Detection (CSMA/CD, Acceso múltiple con escucha de portadora y detección de colisiones).

² Originalmente había un solo protocolo, TCP, que más tarde se escindió en dos: TCP e IP, en 1980.

El segundo elemento creativo fue el uso de ordenadores especiales llamados puertas de acceso o *gateways* a modo de interfaz entre redes diferentes (Cerf 1979). En la actualidad se conocen como *router* y, tal y como su nombre indica, su función es determinar la ruta que los paquetes deben seguir para pasar de una red a otra. Una red dirige paquetes no locales a una puerta de acceso cercana, la cual a su vez los envía a su red de destino. Al repartir la tarea de enrutamiento entre distintas redes y *gateways*, esta arquitectura hacía más fácil el crecimiento de Internet: las redes individuales ya no necesitaban conocer la topología de toda la red, tan sólo saber cómo llegar a la puerta de acceso más cercana; a su vez, las puertas de acceso no precisaban ser capaces de llegar a todas las redes dentro de Internet, únicamente a los anfitriones individuales dentro de una red.

Otro invento notable que contribuyó a hacer más manejable el crecimiento a escala global de Internet fue el Sistema de nombres por dominio o *Domain Name System*, creado en 1985 por Paul Mockapetris (Cerf 1993; Leiner et al. 1997). Uno de los retos de la comunicación en una red de gran tamaño es que por lo general pasa por conocer la dirección del ordenador de destino. Mientras que los seres humanos suelen referirse a los ordenadores con nombres (como «Darpa»), los ordenadores que forman parte de la red se identifican unos a otros mediante direcciones numéricas. En la ARPANET originaria los nombres y direcciones de todos los ordenadores, excepto los anfitriones, se guardaban en un gran archivo que tenía que ser constantemente actualizado y distribuido a todos los anfitriones. Claramente este mecanismo no estaba a la altura de una red de miles e incluso millones de ordenadores. El Sistema de nombres por dominio descentralizó la tarea de localizar direcciones creando grupos de nombres conocidos como *dominios* (tales como .com o .org) y ordenadores especiales llamados *servidores de nombres*, encargados de mantener bases de datos con las direcciones correspondientes a cada nombre de dominio. Para encontrar una dirección el anfitrión únicamente debía introducir sus términos de búsqueda en el servidor apropiado. El nuevo sistema también permitía descentralizar la autoridad para asignar nombres de manera que, por ejemplo, cada país pudiera controlar su propio dominio.

La World Wide Web y otras aplicaciones

La arquitectura de Internet hizo posible construir una infraestructura mundial de datos, pero no se ocupaba directamente del asunto de los contenidos. En la década de 1980 casi todos los contenidos de Internet eran puro texto. Era relativamente difícil para los usuarios localizar la información deseada; para ello debían conocer de antemano la dirección del sitio que contenía los datos, puesto que no existían motores de búsqueda ni vínculos entre distintos sitios. La gran innovación que vino a transformar la manera en que los contenidos de Internet son creados, presentados y encontrados fue la World Wide Web.

La World Wide Web fue la creación de Tim Berners-Lee, un investigador británico del CERN, el laboratorio internacional de Física de Ginebra. Éste imaginó el futuro de Internet como un espacio colaborativo en el que las personas pudieran compartir información de todo tipo. En el modelo que propuso, los usuarios podían crear páginas de contenidos en ordenadores llamados *servidores web*, y las páginas *web* podían verse mediante un programa llamado *browser*. La Web tendría capacidad para manejar información multimedia además de texto y las páginas *web* podrían conectarse entre sí mediante *hipervínculos*, de forma que los usuarios pudieran navegar entre sitios basándose en contenidos relacionados entre ideas de diferentes páginas. Esto crearía una red de conexiones basadas en contenidos en lugar de en infraestructura. Berners-Lee formuló sus ideas en 1989 y en 1990 creó, junto a su colaborador Robert Cailliau, la primera versión operativa de la Web. Los soportes técnicos del nuevo sistema incluían *html* (lenguaje de marcas de hipertexto, empleado para crear páginas *web*), *http* (protocolo de transferencia de hipertexto, para transmitir datos de las páginas *web*) y el *url* (localizador uniforme de recursos, una forma de indicar la dirección de una página *web* determinada).

El uso de la Web se popularizó entre los físicos del CERN, quienes lo extendieron a otros sitios de investigación. En uno de ellos, el US National Center for Supercomputer Applications (NCSA), Marc Andreessen capitaneó el desarrollo de un buscador mejorado llamado Mosaic en 1993. El Mosaic funcionaba tanto en ordenadores personales como en aparatos más grandes y el NCSA lo cargó de forma gratuita en Internet, lo que provocó un interés creciente en la Web. En 1994 se calculaba que había al menos un millón de copias de Mosaic en uso (Schatz y Hardin 1994).

Los hipervínculos de la Web se diseñaron para resolver un persistente problema para los usuarios de Internet: cómo encontrar información dentro de un sistema tan grande. Con el fin de resolverlo, en la década de 1990 se desarrollaron varias herramientas de búsqueda. Una de las primeras fue Archie (1990), el cual enviaba preguntas a ordenadores conectados a Internet y reunía listados a partir de ficheros de libre acceso. Gopher (1991) era un sistema de listado específico para la Web, mientras que Yahoo (1994) era un directorio de páginas *web* organizadas por temas. El personal de Yahoo categorizaba dichas páginas de forma manual; sin embargo, dado el gran volumen de información que se iba acumulando en la Web, surgieron una serie de servicios que tenían por objeto automatizar las búsquedas. El más exitoso de estos *motores de búsqueda* fue Google (1998). Estas herramientas transformaron la forma en que los usuarios localizaban información en la Web, al permitirles buscar información sobre un tema en particular en un gran número de fuentes, en lugar de tener que conocer de antemano cuáles de éstas podrían contener información relevante.

Al igual que Internet, la Web se diseñó para ser flexible, expandible y descentralizada, de forma que invitara al público a desarrollar nuevas maneras de usarla. La generalización del uso de la World Wide Web coincidió con el traspaso, en 1995, de la gestión de Internet del gobierno al sector privado. Ello hizo desaparecer numerosas barreras para el uso comercial de Internet y trajo consigo el auge de las empresas «punto com» en la década de 1990, cuando se invirtieron grandes sumas de capital en proyectos de comercio electrónico. Aunque la burbuja de las «punto com» estalló en el año 2000, el fenómeno sirvió para fomentar un conocimiento de Internet como motor económico y no una mera innovación tecnológica a nivel popular. La llegada del siglo *xxi* también trajo consigo la proliferación de los *medios de comunicación sociales*, que proporcionaban nuevas formas de relacionarse y compartir información en línea. Éstos incluían *weblogs* (1997), *wikis* (1995), *file sharing* (compartir archivos, 1999), *podcasting* (2004) y sitios de redes sociales, así como una gran variedad de juegos multijugador.

Internet y la sociedad: éxitos y desafíos a futuro

Después de un siglo y medio de investigación e innovaciones, Internet quedó firmemente establecido como un recurso accesible para muchas personas y que ofrecía una amplia variedad de beneficios potenciales. Los usuarios tenían mayor acceso a todo tipo de información y los gobiernos y negocios disponían de una nueva plataforma para proporcionar información y servicios. El comercio electrónico trajo consigo crecimiento económico, mayor amplitud de oferta para los consumidores y oportunidades para productores, que trabajaban en las áreas más desfavorecidas, de llegar a nuevos mercados. Una variedad de opciones de comunicación, desde el correo electrónico a elaboradas redes sociales, hacían más fácil a amigos y familias separados geográficamente estar en contacto, y a extraños formar «comunidades virtuales» en torno a intereses comunes. Los organizadores de base adoptaron Internet como plataforma de activismo político y social y empezaron a emplearlo para movilizar respuestas masivas a desastres naturales y violaciones de derechos humanos. Usuarios de todas las edades saludaron a Internet como un medio de expresión personal y las nuevas aplicaciones ayudaron a democratizar la tecnología, posibilitando que el ciudadano de a pie generara y difundiera noticias, información, opinión y entretenimiento.

Sin embargo, llegado el siglo *xxi* son todavía muchos los desafíos a los que se enfrenta Internet. Los usuarios eran víctimas de prácticas abusivas como el *spam* (correo comercial no deseado), virus, usurpación de identidad y piratería. Los expertos técnicos responden con soluciones que intentan minimizar estos peligros continuos diseñando sistemas anti virus, filtros, transacciones seguras y otras medidas. Pero hay otras cuestiones demasiado controvertidas para que una solución técnica satisfaga una opinión pública dividida, especialmente cuando se trata de actividades que traspasan las fronteras nacionales. Algunos

gobiernos optaron por limitar de forma estricta y supervisar las actividades en línea de sus ciudadanos. Cuando las asociaciones pro derechos humanos protestaron calificando dichas medidas de censura y represión, los gobiernos en cuestión defendieron su derecho a proteger la seguridad y la moral públicas. Otros grupos se quejaron de que Internet estaba *demasiado* abierta a contenidos objetables o ilegales, tales como pornografía infantil o la piratería de películas, canciones y *software*. Los mecanismos de filtrado y protección han proporcionado medios para restringir el flujo de esta clase de información, pero son en sí controvertidos. La gestión de Internet también ha resultado ser un asunto espinoso, con muchas naciones exigiendo un funcionamiento más internacional y con menor intervención estadounidense a la hora de gestionar el sistema de direcciones y nombres en Internet.³ Otra cuestión técnica pero con ramificaciones políticas fue la propuesta de transición del antiguo Protocolo de Internet, llamado IPv4 a uno nuevo de nombre IPv6, que proporcionaría un número mucho mayor de direcciones disponibles (Bradner y Mankin 1996); esto fue en parte una respuesta al hecho de que Estados Unidos poseía una porción totalmente desproporcionada de las direcciones del IPv4. El IPv6 se propuso como nuevo estándar de Internet en 1994, pero debido a desavenencias políticas y técnicas, quince años más tarde sólo era usado para una proporción minúscula del tráfico total de la red (DeNardis 2009). A pesar de todos estos obstáculos, el proceso de desarrollo descentralizado y basado en el consenso de Internet continúa funcionando lo suficientemente bien como para mantenerse próspero y en crecimiento y cambio constantes.

Tal vez lo más preocupante sea la persistencia de la desigualdad en acceso a Internet y a las oportunidades que éste conlleva de mejora económica, participación política, transparencia gubernamental y desarrollo científico y tecnológico. Aún existen diferencias significativas entre regiones pobres y ricas, rurales y urbanas, mayores y jóvenes. Naciones Unidas publicó en 2007 un informe según el cual las desigualdades en cuanto a acceso digital eran todavía enormes: «Más de la mitad de la población de las regiones desarrolladas usaba Internet en 2005, comparado con un 9% en las regiones en vías desarrollo y un 1% en los países menos desarrollados» (Naciones Unidas 2007, 32). Para ayudar a paliar estas diferencias, Naciones Unidas y el Sindicato Internacional de Comunicaciones patrocinaron dos cumbres mundiales: en la Information Society de Ginebra en 2003 y en Túnez, en 2005, con el fin de diseñar un plan de acción que proporcione acceso a las tecnologías de la información y la comunicación a todas las gentes del mundo (WSIS 2008). Ingenieros informáticos también dedican sus esfuerzos a hacer Internet más accesible a los desfavorecidos del mundo. Por ejemplo, en 2001 un grupo de ingenieros informáticos indios revirtieron el paradigma de ordenadores personales costosos y de alto consumo de energía creando el Simputer: un ordenador sencillo, de

³ La «internacionalización» de la gestión de Internet fue el tema central de la Cumbre Mundial sobre la Sociedad de la Información patrocinada por Estados Unidos en 2005.

4
Los creadores y administradores del proyecto Simputer fueron Vijay Chandru, Swami Manohar, Ramesh Hariharan, V. Vinay, Vinay Deshpande, Shashank Garg y Mark Mathias (<http://www.simputer.org/simputer/people/trustees.php>).

bajo coste y consumo dotado de un interfaz en varias lenguas y que podía ser compartido por todos los habitantes de un pueblo (Sterling 2001).⁴ De forma similar, Nicholas Negroponte puso en marcha el proyecto «Un ordenador personal por niño» en 2005 para cubrir necesidades educativas en países en vías de desarrollo. Para ayudar a adaptar la tecnología a las necesidades de cada lugar, la famosa diseñadora Mary Lou Jensen inventó una pantalla muy práctica y económica legible a la luz del sol, y el diseñador de *software* Walter Brender hizo lo mismo con una interfaz de usuario intuitiva (One Laptop Per Child 2008; Roush 2008). El Desafío de Estocolmo, un evento que se celebra anualmente desde 1995, sirve de escaparate a cientos de iniciativas innovadoras procedentes de todo el mundo que utilizan tecnologías de la comunicación e información para promover el desarrollo (Desafío de Estocolmo 2008).

Expandir las fronteras de Internet ya no es patrimonio exclusivo de los ingenieros informáticos y científicos, sino que implica cada vez en mayor grado la colaboración entre organizaciones civiles, gobiernos y ciudadanos. Los valores que guían el desarrollo técnico y social de Internet han sido complementarios: mayor acceso, apertura a la diversidad, descentralización, toma de decisiones por consenso entre un número elevado de participantes y la posibilidad de que los usuarios contribuyan de forma activa. Desde el punto de vista técnico estos objetivos se han alcanzado por medio de una arquitectura de capas, protocolos abiertos y un proceso colaborativo a la hora de aprobar y diseñar cambios, mientras que las metas sociales han requerido del liderazgo de gobiernos y de la inspiración de personas que fueron capaces de ver el potencial de Internet para la comunicación, la cooperación y la expresión individual.

Bibliografía

- Abbate, J. *Inventing the Internet*. Cambridge: MIT Press, 1999.
- Abramson, N. «The ALOHA System—Another Alternative for Computer Communications». *Proceedings, AFIPS Fall Joint Computer Conference*. Montvale, Nueva Jersey: AFIPS Press, 1970: 281–285.
- Baran, P. *On Distributed Communications*. Santa Monica, CA: RAND Corporation, 1964.
- Berners-Lee, T. *Weaving the Web*. Nueva York: Harper Collins, 1999.
- Bradner, S. «The Internet Standards Process—Revision 3». Network Working Group Request for Comments 2026, octubre 1996. Disponible en Internet en: <http://www.rfc-editor.org/>
- , y A. Mankin. «The Recommendation for the IP Next Generation Protocol». Network Working Group Request for Comments 1752, enero 1995. Disponible en Internet en: <http://www.rfc-editor.org/>
- Campbell-Kelly, M. y W. Aspray. *Computer: A History of the Information Machine*. Nueva York: Basic Books, 1996.
- Cerf, V. G. «DARPA Activities in Packet Network Interconnection». En K. G. Beauchamp, ed. *Interlinking of Computer Networks*. Dordrecht, Holland: D. Reidel, 1979.
- , «Oral history interview» por Judy O'Neill (Reston, VA, April 24, 1990), OH 191. Minneapolis, Minnesota: The Charles Babbage Institute, University of Minnesota, 1990. Disponible en Internet en: <http://www.cbi.umn.edu/oh/display.phtml?id=118>.
- , «How the Internet Came to Be». En B. Aboba, ed. *The Online User's Encyclopedia*. Addison-Wesley, 1993.
- , y R. E. Kahn. «A Protocol for Packet Network Intercommunication». *IEEE Transactions on Communications* COM-22 (mayo 1974): 637–648.
- Cumbre mundial sobre la sociedad de la información. <http://www.itu.int/wsis/> (consultado el 5 de septiembre de 2008).
- DeNardis, L. *Protocol Politics: The Globalization of Internet Governance*. Cambridge: MIT Press, 2009.
- Desafío de Estocolmo. www.challenge.stockholm.se/ (consultado el 5 de septiembre de 2008).
- Laws, J. y V. Hathway. «Experience From Two Forms of Inter-Network Connection». En K. G. Beauchamp, ed. *Interlinking of Computer Networks*. OTAN, 1978: 273–284.
- Leiner, B. M., V. G. Cerf, D. D. Clark, R. E. Kahn, L. Kleinrock, D. C. Lynch, J. Postel, L. G. Roberts y S. Wolff. «A Brief History of the Internet». 1997. Disponible en Internet en: <http://www.isoc.org/internet-history>. Revisado en febrero de 1997.
- McGrath, D. «Minitel: The Old New Thing». *Wired*, 18 de enero de 2004. Disponible en Internet en: <http://www.wired.com/science/discoveries/news/2001/04/42943>.
- Metcalfe, R. M. *Packet Communication*. San José: Peer-to-Peer Communications, 1996.
- Naciones Unidas. *The Millennium Development Goals Report 2007*. Nueva York: Naciones Unidas, 2007.
- One Laptop Per Child. <http://www.laptop.org/> (consultado el 5 de septiembre de 2008).
- Pouzin, L. «Presentation and Major Design Aspects of the CYCLADES Computer Network». En R. L. Grimsdale y F. F. Kuo, eds. *Computer Communication Networks*. Leyden: Noordhoff, 1975.
- Quarterman, J. S. *The Matrix: Computer Networks and Conferencing Systems Worldwide*. Burlington, Massachusetts: Digital Press, 1990.
- Roush, W. «One Laptop Per Child Foundation No Longer a Disruptive Force, Bender Fears». *Xconomy*, 24 de abril de 2008. Disponible en Internet en: <http://www.xconomy.com/boston/2008/04/24/>
- Schatz, B. R. y J. B. Hardin. «NCSA Mosaic and the World Wide Web: Global Hypermedia Protocols for the Internet». *Science* 265 (1994): 895–901.
- SITA. «SITA's history and milestones». http://www.sita.aero/News_Centre/Corporate_profile/History/ (actualizado el 21 de julio de 2006, consultado el 5 de septiembre de 2008).
- Sterling, B. «The Year in Ideas: A to Z; Simputer». *The New York Times*, 9 de diciembre de 2001.

el siglo del gen. biología molecular y genética

GINÉS MORATA

En el siglo xx la sociedad humana ha incorporado de forma masiva el desarrollo tecnológico. Durante gran parte del siglo las mayores contribuciones tecnológicas han venido de los derivados de las ciencias físicas: automóviles, teléfono, aviones, plásticos, ordenadores, etc. La introducción de estos factores ha modificado la sociedad y el comportamiento humano más que los propios acontecimientos políticos y sociales. Sin embargo, durante la segunda mitad del siglo xx, y especialmente en las últimas dos décadas, han aparecido unas tecnologías biológicas que poseen un enorme potencial médico y social. Estas nuevas tecnologías, extraordinariamente poderosas, son las derivadas del progreso explosivo de la biología molecular en la segunda mitad del siglo xx. Ofrecen una nueva imagen de la evolución de la vida en el planeta y están llamadas a revolucionar la propia estructura de la sociedad humana.

Quizá la persona que más lúcidamente ha profundizado en estas ideas ha sido Sydney Brenner, uno de los científicos más brillantes del siglo xx que pasará a la historia de la ciencia por sus enormes aportaciones a la biología molecular, ciencia que contribuyó decisivamente a crear. Afirma Brenner que la nueva biología nos acerca a la comprensión de nosotros mismos, al entendimiento de los humanos como organismos: «Por primera vez podemos plantear el problema fundamental del hombre y empezar a comprender nuestra evolución, nuestra historia, nuestra cultura y nuestra biología como un todo».

En esta exposición trataré de la historia de los acontecimientos científicos que han dado lugar a esta situación y especularé brevemente sobre las implicaciones que tienen estos nuevos descubrimientos en la sociedad del futuro, e incluso en nuestra propia comprensión de la naturaleza humana.

Los hitos del conocimiento biológico

A lo largo de la historia de la biología han existido tres grandes revoluciones, entendiendo el concepto revolución como la introducción de un descubrimiento que es importante por sí mismo y que además da lugar a un cambio radical en el enfoque que hasta ese momento se daba a la disciplina. No sólo se trata del significado de la nueva información, sino también del efecto que tiene en el enfoque general de disciplina.

La primera revolución se produjo en 1860 con las teorías evolucionistas de Darwin y Wallace, que defendían la universalidad del origen de los seres vivos. La segunda revolución fue el descubrimiento de la universalidad del mecanismo de la información biológica propuesta por Watson y Crick en 1953. La tercera revolución ha sido el descubrimiento de la universalidad del diseño animal y de los procesos básicos de regulación de las funciones biológicas. Esta última revolución ha sucedido entre los años 1985-2000 del pasado siglo y, a diferencia de las anteriores, es el resultado de las contribuciones de un grupo relativamente numeroso de investigadores. Estos tres hechos

han conducido a una nueva comprensión del fenómeno evolutivo y de la propia biología de los seres humanos.

El hecho evolutivo

La idea de que las especies cambian con el tiempo es muy antigua y ciertamente anterior a la propuesta darwinista. En el año 520 ANE, Anaximandro de Mileto, en su tratado *On Nature*, introdujo la idea de la evolución y que la vida empezó en los océanos. John Ray en su libro *Historia Plantarum*, publicado en 1686, catalogó 18.600 tipos de plantas y propuso la primera definición de especie basada en una descendencia común. El propio abuelo de Darwin, Erasmus Darwin, propuso explícitamente que las especies animales cambian con el tiempo.

Lo que diferencia a Darwin y Wallace de sus predecesores es que propusieron un mecanismo evolutivo *plausible* basado en la idea de selección natural. Darwin, en particular, proponía que en la supervivencia de los individuos especialmente favorecidos estaba la fuerza de la selección natural ya que su mayor capacidad de supervivencia les dotaba también de una mayor capacidad de transmitir sus características a su descendencia. Mediante este proceso las características de las poblaciones de cada especie en particular se modificaban gradualmente a través de sucesivas generaciones.

Darwin disponía además de información que sus predecesores no conocían y ésta ayudó mucho a la comprensión del fenómeno evolutivo. Por ejemplo, se sabía que la edad de la Tierra era mucho mayor que lo que se había supuesto, lo cual concedía mucho más tiempo para el cambio gradual que preconizaba la teoría de la selección natural. También existía un registro fósil ya muy elaborado que permitía comprobar la existencia de un cambio gradual en muchas líneas de animales y plantas, lo cual apoyaba claramente la propuesta darwiniana. Se sabía también que la selección artificial es capaz de generar cambios morfológicos muy profundos en poco tiempo. Esto es evidente si, por ejemplo, consideramos la gran cantidad de razas de perros que existen. Todas derivan del lobo, pero a través de 5.000 o 10.000 años de evolución artificial, no natural, el hombre ha conseguido crear una gran diversidad de razas caninas, lo que indica hasta qué punto es versátil el material biológico cuando se somete a selección.

Si tuviéramos que resumir las implicaciones de la teoría evolutiva podríamos centrarnos en tres puntos: 1) todos los seres vivos tienen un origen común, 2) ha habido un proceso de cambio gradual que a lo largo de muchos millones de años ha dado lugar a toda la diversidad biológica del planeta y, por último, 3) la especie humana es simplemente una más de los cientos de millones de especies que existen o han existido. La propuesta darwiniana reflejaba un cambio copernicano en lo que respecta a la posición de la especie humana como entidad biológica; el hombre dejaba de ser el centro de la creación para convertirse en simplemente una especie más de los millones de especies creadas por la evolución. No es sorprendente que se produjera una

gran reacción social en la época. Aún en la época actual no es aceptada por muchos miembros de la sociedad. Según el Instituto Gallup, en 2004 más de la mitad de los norteamericanos creían que el hombre ha sido creado literalmente tal como dice la Biblia, hace unos 10.000 años.

Genética y evolución:

definición operacional del gen

Darwin proporcionó una explicación descriptiva, plausible, pero no mecanística de la diversidad biológica. El asunto es: si todos los organismos vivos tienen un origen común, ¿qué función biológica es común para todos, se transmite de padres a hijos y es modificable para generar la diversidad biológica? En su tiempo, Darwin no podía responder estas cuestiones.

Fue precisamente el planteamiento de estas cuestiones lo que dio origen a la genética, la disciplina que estudia cómo se transmite y modifica la información biológica. La primera señal de la existencia de información genética heredable se debe a Gregor Mendel, un monje agustino que demostró que la forma o el color de los guisantes se transmite fielmente de una generación a otra.

Sin embargo, el progreso de la genética en el siglo xx se debe en gran parte a la mosca del vinagre, *Drosophila melanogaster*, un organismo que se ha convertido en el objeto de estudio clásico de investigación en genética, ya que se cría fácilmente en el laboratorio, su ciclo biológico es muy corto –lo que resulta muy útil para estudiar la herencia de los diversos caracteres de una generación a otra– y resulta totalmente inocua para los humanos. Los estudios en *Drosophila* permitieron identificar muchos caracteres heredables concretos –genes–, demostraron que están localizados y alineados en el núcleo de las células en unos orgánulos llamados cromosomas y que cada gen está situado en una posición específica del cromosoma. También demostraron que en la naturaleza aparecen variaciones heredables –mutaciones– de los genes y que estas mutaciones son la fuente de variación biológica necesaria para el proceso evolutivo. Estas mutaciones también se pueden inducir de forma artificial mediante irradiaciones o el uso de compuestos químicos. En conjunto, lo que la genética de *Drosophila* descubrió es que el motor real de la evolución son los genes, que constituyen la información genética heredable y que son modificables.

Después de más de un siglo de investigaciones en esta mosca, el conocimiento de su genética es el más completo del Reino Animal y se han desarrollado unos conceptos y tecnologías que permiten realizar experimentos que no se pueden hacer en ninguna otra especie.

Naturaleza de la información genética

El problema que surgió a continuación, aproximadamente en los años cuarenta del pasado siglo, era conocer la naturaleza física del gen; cuál era su composición química.

La solución a este problema dio lugar a lo que yo llamo la segunda revolución de la biología: la dilucidación por

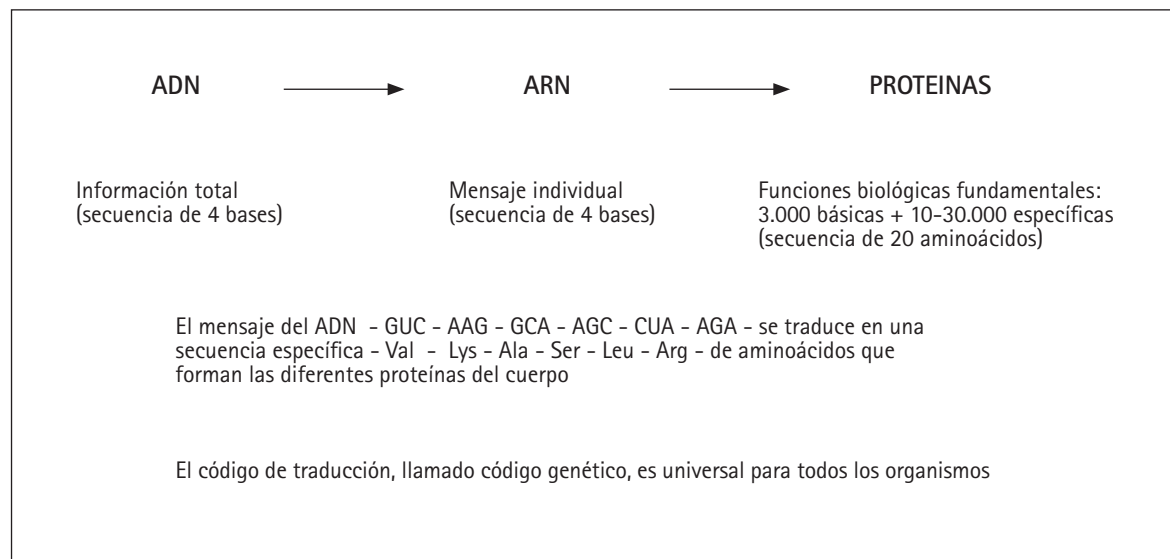


Figura 1. Traducción del texto genético.

Watson y Crick de la naturaleza y estructura de la información genética, el ADN. El famoso artículo publicado en la revista *Nature* el año 1953 fue el comienzo de una revolución biológica destinada a cambiar el propio rumbo de la humanidad. El ADN es una molécula con estructura de doble hélice formada por dos largas cadenas de moléculas de un azúcar -deoxi-ribosa- unidas por fosfatos. Conectando ambas cadenas, como peldaños de una escalera, otras moléculas llamadas bases nitrogenadas mantienen la estabilidad de la estructura. Como notaron inmediatamente Watson y Crick, la propia estructura de la molécula explica el mecanismo de replicación dando lugar a moléculas iguales y, por lo tanto, asegurando la fidelidad de la información biológica a través de generaciones.

Pero además, la estructura del ADN indicaba que la información biológica radicaba en la secuencia a lo largo de la molécula de las cuatro bases nitrogenadas llamadas timina (T), guanina (G), adenina (A) y citosina (C). Lo que un organismo hereda de sus progenitores y que va a condicionar sus características biológicas es simplemente una secuencia escrita en un lenguaje de cuatro letras.

El descubrimiento de la estructura y función del ADN modificó el enfoque experimental de la biología: todos los organismos están cifrados en un lenguaje de cuatro letras, A, T, C y G. A partir de entonces la biología se centró en el estudio del ADN, sus propiedades y su estructura. La primera secuencia completa de ADN que se obtuvo de un organismo, el bacteriofago ØX174, contiene 5.000 letras -llamadas bases-. Por comparación la secuencia de ADN de un gusano nemátodo consta de 90 millones de pares de bases; la secuencia de la mosca del vinagre *Drosophila* consta de 120 millones de pares de bases y la del ser humano consta de 3.300 millones de pares de bases. Cada una de estas secuencias representa una especie de fórmula para construir la especie en cuestión.

Un código genético universal

El problema es que los procesos vitales no están catalizados por el ADN, sino por las proteínas; el ADN es simplemente una receta que ha de ser traducida en toda la variedad de proteínas, unas 3.000 básicas, que se encargan de las funciones vitales, entre ellas de la propia replicación y expresión del ADN.

Las proteínas están constituidas por combinaciones de 20 aminoácidos, de forma que cada proteína es diferente de las demás debido a que está formada por una secuencia específica de aminoácidos. Así pues, hay que traducir la secuencia de cuatro bases heredada de los progenitores en secuencias de 20 aminoácidos para producir las proteínas que son el sostén de las funciones biológicas. El desciframiento del código de traducción, el código genético, fue uno de los grandes primeros éxitos de la biología molecular. Los laboratorios de Ochoa, Nirenberg y Brenner fueron decisivos para descifrar el mecanismo de traducción. Estos investigadores demostraron que cada aminoácido está codificado por una secuencia específica de tres bases -tripleto-, asegurando de ese modo que cada gen, que es una secuencia particular del ADN total, se traduce en una proteína específica. El tripleta AAG codifica para el aminoácido lisina, mientras que GCA codifica alanina y AGA arginina. De esta forma el ADN de secuencia AAGGCAAGA se traduciría en la secuencia de aminoácidos lisina-alanina-arginina (figura 1).

Lo interesante del código genético es que es universal para todos los organismos. La universalidad del código es, en sí misma, una prueba de la evolución. Todos los organismos tenemos el mismo código genético simplemente porque todos lo hemos heredado de un antepasado ancestral.

En este contexto un gen es simplemente una secuencia concreta de ADN que codifica para una proteína específica encargada de una función concreta, por ejemplo la hemoglobina necesaria para la respiración o la miosina del músculo.

Desarrollo de la biología molecular

El descubrimiento de que el ADN es el manual de instrucciones para hacer un ser vivo y el desciframiento de los mecanismos básicos de la función génica, el código genético y la manufactura de proteínas, marcan el comienzo de la biología molecular. El estudio del ADN, su estructura y sus propiedades se convirtió en el principal foco de esta disciplina a partir de los años setenta del pasado siglo. Esta concentración de esfuerzos ha dado lugar a conceptos y técnicas extraordinariamente poderosas que permiten manipular el ADN con gran eficiencia. Son estas técnicas las que permiten el clonaje de genes, la generación de animales y plantas transgénicas, la posibilidad de terapia génica y los Proyectos Genoma. La generación de organismos transgénicos, esto es, organismos a los que se les ha introducido genes de otra especie, se deriva del hecho de que todos los ADN, de cualquier origen, son químicamente idénticos y que un gen es simplemente un fragmento de ADN. Esto permite mezclar por métodos químicos fragmentos de ADN —genes— de origen heterólogo. Una vez que se han desarrollado métodos para introducir estos fragmentos en el organismo receptor, éste dispone ahora de un gen de origen diferente. Un ejemplo claro lo constituyen, por ejemplo, las estirpes de levaduras a las que se les ha introducido el gen humano que codifica para la insulina. Mediante este procedimiento las levaduras transgénicas manufacturan insulina humana.

El gran desarrollo de estos procedimientos en años recientes ha permitido generar plantas —trigo, soja, arroz, etc., que ya están en el mercado— y animales transgénicos de muchas especies, ratas, ratones, cerdos, moscas, etc. Es importante señalar que los métodos que se utilizan para las varias especies animales son muy parecidos y constituyen la base de aplicaciones para su uso terapéutico en la especie humana, con objeto de curar enfermedades genéticas

mediante terapia génica. En el año 2000 se publicó en la revista *Science* el primer ensayo de terapia génica mediante el que se logró curar una inmunodeficiencia severa a varios niños. Desgraciadamente, estos ensayos se tuvieron que interrumpir debido a efectos perjudiciales del procedimiento; tres de los niños curados desarrollaron posteriormente un cáncer. Este ejemplo ilustra, al mismo tiempo, el potencial de estos nuevos métodos y el hecho de que todavía están en una fase muy temprana de su desarrollo. Dada la velocidad con la que se está progresando es de esperar que estén disponibles en un futuro no demasiado lejano.

El diseño genético del cuerpo animal

Uno de los aspectos en que la biología molecular ha progresado de forma importante y con considerables aplicaciones en lo que respecta a la biología humana es en el área del diseño genético del cuerpo de los animales.

Inicialmente, los experimentos en biología molecular utilizaban organismos unicelulares, bacterias o virus para el estudio de las propiedades y función del ADN. Estos estudios produjeron resultados muy importantes, como se ha descrito anteriormente, pero por su propia naturaleza no permitían conclusiones sobre cuál es el control genético del desarrollo de los organismos complejos, como una mosca o un ratón, formados por asociaciones de células que han de agruparse de forma correcta en una estructura tridimensional.

Tomemos como ejemplo una mariposa (figura 2); cada célula individual tiene que realizar las funciones biológicas primarias, síntesis de proteínas, replicación del ADN, etc., pero además tiene que disponerse, agruparse con otras y diferenciarse para hacer órganos específicos, ojos, alas, patas etc., los cuales han de ensamblarse con los demás órganos de forma que cada uno de ellos aparezca en el lugar adecuado. En el diseño de un animal hay que disponer las diversas partes del cuerpo en las tres dimensiones del espacio; los ejes antero-posterior, dorso-ventral y próximo-distal. Este problema del diseño corporal ha sido uno de los grandes retos de la genética de los organismos superiores: cómo los genes especifican la información posicional de las diversas partes del cuerpo de forma que las células que van a hacer ojo saben que han de disponerse en la parte anterior del cuerpo y las que van a formar la patas deben de estar en la parte ventral. En otras palabras, ¿cuál es la descripción genética de un organismo tridimensional? En un insecto como una mariposa distinguimos morfológicamente una parte cefálica, una parte torácica y una parte abdominal, pero no hay garantía de que esta descripción corresponda a la verdadera descripción genética del organismo. Es en este asunto de la descripción genética de los animales donde se ha progresado de forma notable en los últimos treinta años.

Las claves del diseño genético del cuerpo animal están en los llamados genes homeóticos, ahora llamados Hox. Estos forman una maquinaria genética que ha sido estudiada con gran detalle en la mosca del vinagre *Drosophi-*

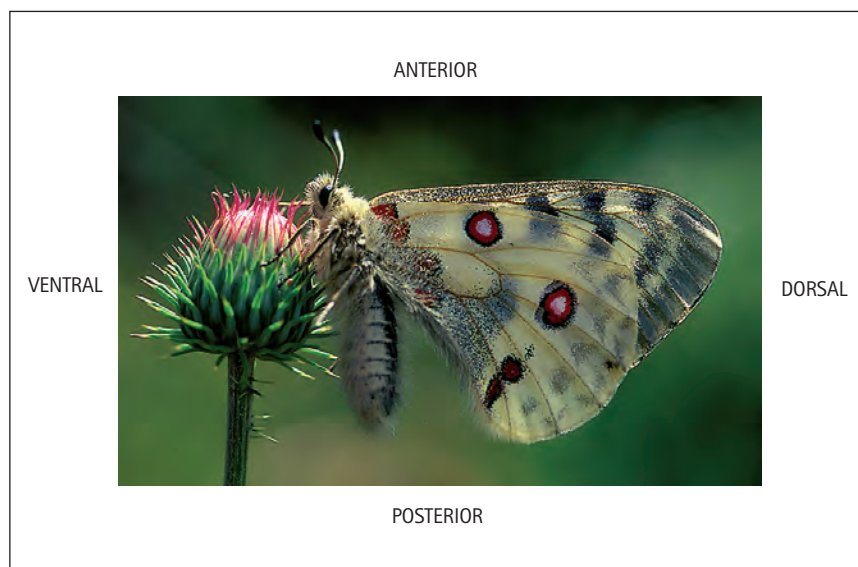


Figura 2. Ejes del cuerpo animal.

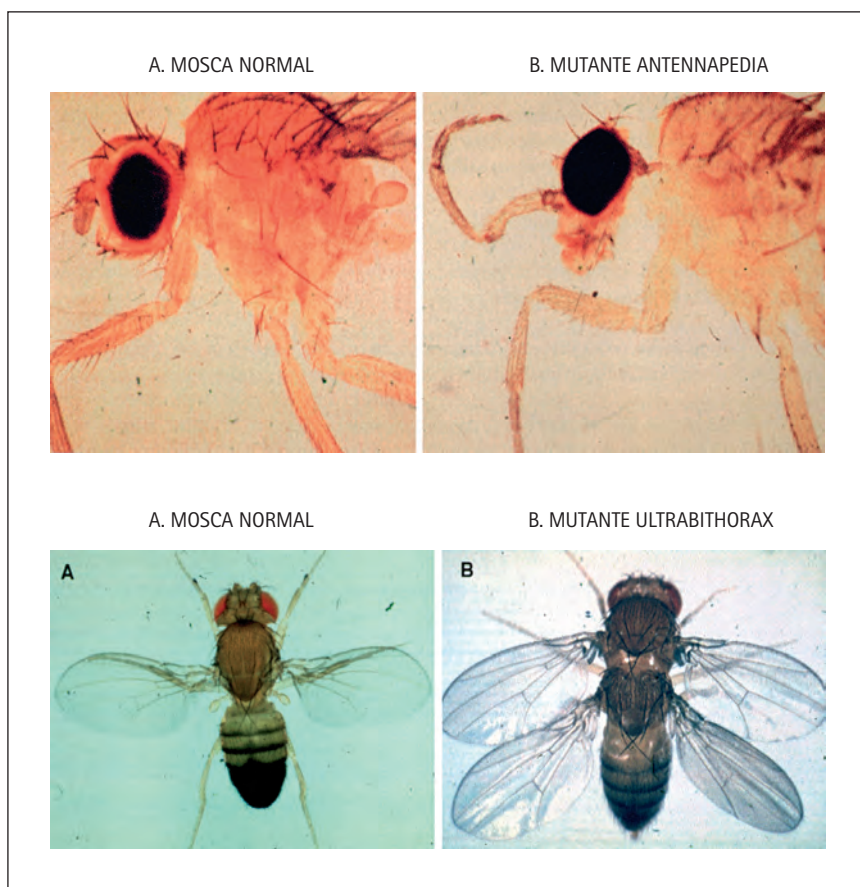


Figura 3. Mutaciones homeóticas de *Drosophila*.

la. Lo característico de estos genes es que sus mutaciones transforman unas partes del cuerpo en otras (figura 3). Una mutación como *Antennapedia* (*Antp*) por ejemplo transforma la antena en pata, o una mutación como *Ultrabithorax* (*Ubx*) transforma el apéndice halterio en ala, dando lugar a una mosca con cuatro alas. Lo interesante de estas transformaciones es que, a pesar de que la construcción general del cuerpo es errónea, la morfología de las partes es normal: la pata que aparece en la antena de *Antp* es normal, lo que es anómalo es el sitio donde aparece. Igualmente, las alas transformadas que aparecen en la mosca *Ubx* tienen morfología y tamaño normal de alas. Lo único anormal es el sitio donde aparecen. La implicación de estos fenotipos es que lo que controlan los genes Hox no es la morfología de las estructuras sino el diseño general del cuerpo, la información posicional a la que me refería anteriormente, que hace que cada órgano aparezca en el lugar correspondiente.

Los genes homeóticos son pues genes reguladores de alto rango que determinan el tipo de desarrollo de las diversas partes del cuerpo de *Drosophila*. Una pregunta muy importante que se hizo en los años ochenta del siglo pasado fue cuántos genes homeóticos existen. La identificación de todos ellos era de esperar que permitiera dilucidar la lógica genética que subyace en el diseño del cuerpo.

Estudios realizados en Estados Unidos y en España demostraron que, sorprendentemente, el número de genes Hox es pequeño. En la *Drosophila* solamente existen nueve genes Hox que establecen las coordenadas espaciales en el eje antero-posterior, reconocen los valores posicionales en este eje y determinan la adquisición del programa de desarrollo adecuado para generar la parte correspondiente del cuerpo. Estos resultados eran ciertamente interesantes pero se referían a la mosca del vinagre; en principio no se sospechaba que pudieran tener un valor general para explicar el diseño del cuerpo de otros animales, incluyendo la especie humana.

Sin embargo, el progreso de la biología molecular en los años setenta-ochenta del pasado siglo permitió el aislamiento molecular —clonaje— y la secuenciación de los genes Hox de *Drosophila*. A finales de 1985 todos estos genes ya había sido clonados y secuenciados. Un descubrimiento extraordinariamente importante que se realizó cuando se compararon sus secuencias es que todos estos genes contenían una secuencia en común, que se llamó homeobox. Las implicaciones del descubrimiento de la homeobox fueron muy importantes: 1) Esta secuencia codifica para un motivo de unión al ADN, indica que las proteínas homeóticas funcionan como factores de transcripción y regulan la actividad de otros genes subsidiarios, 2) la presencia de la misma secuencia en todos los genes Hox indica que estos genes tienen un origen común y 3) la secuencia homeobox es un marcador molecular de los genes Hox que permitió identificarlos en organismos —la especie humana, por ejemplo— en los que es imposible detectarlos por procedimientos genéticos convencionales. Como veremos a continuación este último aspecto resultó ser de gran importancia.

Un diseño genético universal

El hecho de que la homeobox sea un marcador molecular de los genes Hox permitió identificar el complejo Hox en muchos grupos del Reino Animal, e hizo de estos genes el objeto fundamental de estudio de la biología del desarrollo durante los años ochenta y hasta principios de los noventa. El resultado general es que el complejo Hox se ha encontrado en todos los grupos animales en los que se ha buscado. Es, pues, una característica universal del genoma de todos los animales, incluyendo a la especie humana. Los humanos tenemos un complejo Hox muy parecido al de *Drosophila*, sólo que en vez de tener una copia por genoma tenemos cuatro.

Los estudios de *Drosophila* habían establecido previamente que la función de estos genes era determinar el desarrollo de las diversas partes del cuerpo, pero no existía evidencia de qué función realizan en otros organismos. La dificultad para estudiar este aspecto es que el análisis genético realizado en *Drosophila* no es posible en muchos vertebrados y totalmente imposible en la especie humana, por lo tanto hubo que recurrir a otros métodos.

Las tecnologías moleculares que se desarrollaron en los años ochenta y noventa permiten generar individuos, moscas *Drosophila* en este caso, a los que se les puede introducir un gen de otra especie y estudiar en este sistema heterólogo su función. Diversos experimentos de este tipo han permitido concluir que los genes Hox humanos y de otros vertebrados funcionan de forma parecida o igual a sus homólogos de *Drosophila*. La conservación funcional llega al extremo de que genes humanos o de ratón son capaces de reemplazar a sus homólogos de *Drosophila*, éste es el caso del gen de ratón Hoxd13 que si se introduce en la mosca es capaz de programar el desarrollo de la parte posterior de *Drosophila* como el propio gen de la mosca. Otros ejemplos muy llamativos son, por ejemplo, genes de *Drosophila* *apterous* y *eyeless* que tienen homólogos conocidos en humanos. *Apterous* es necesario para hacer alas; las mutaciones de este gen producen individuos sin alas. *Eyeless* se necesita para programar el desarrollo del ojo; los mutantes en este gen no tienen ojos.

Cuando a una mosca mutante *apterous* se le introduce el gen humano es capaz de formar alas de mosca. Aunque los humanos no tengamos alas de mosca, tenemos un gen capaz de reemplazar al de *Drosophila* que programa la formación de las alas de mosca gracias a un mecanismo de conservación funcional.

Igualmente, el gen de ratón homólogo de *eyeless*, llamado *small eye*, es capaz de inducir ojos de mosca (figura 4). Experimentos similares con genes de otros organismos han llevado a la conclusión de que el diseño genético de los ojos de todos los animales, moscas, pulpos, seres humanos, es el mismo. El invento evolutivo de un órgano

receptor de luz conectado al cerebro tuvo lugar hace 540 millones de años y ha sido heredado por todos los organismos multicelulares. Estos experimentos ilustran un principio general del fenómeno evolutivo: cuando aparece un mecanismo que opera adecuadamente, la programación genética de este mecanismo queda fijada en el genoma y, a partir de ahí, no se modifica o se modifica muy poco.

La conclusión general de todo lo anterior es que el mecanismo general de diseño genético de los animales, basado en los genes Hox y derivados, es común para todo el Reino Animal. Seguramente la explosión del Cámbrico, esto es, la aparición repentina de los Bilateralia con los órganos dispuestos a lo largo de los tres ejes espaciales, es el resultado de la aparición en el Cámbrico Inferior del complejo Hox y sus derivados. La similitud en secuencia de los genes del complejo indica que derivan de un gen ancestral que sufrió varias duplicaciones en tándem, originándose así el conjunto de genes ligados que forman el conjunto.

Por tanto, podemos afirmar que todos los seres vivos comparten las mismas funciones biológicas básicas. De todos estos estudios ha surgido una visión unificadora de los procesos biológicos basada, en última instancia, en el proceso evolutivo. Los organismos tienen un origen común, como propusieron Darwin y Wallace, comparten el mecanismo de almacenamiento y liberación de la información genética basado en la universalidad de la función del ADN, del ARN y en el mecanismo de código genético. Por último, todos los componentes del Reino Animal comparten el mismo proceso genético de diseño corporal.

Una implicación de importancia que deriva de estas observaciones es que muchos de los aspectos del desarrollo del cuerpo humano se pueden estudiar en organismos modelo, moscas, gusanos, ratones, en el entendimiento de que la base genética/molecular de estos procesos es común para todas las especies y, por lo tanto, muchos de los procesos involucrados también lo serán. Un ejemplo típico de este enfoque lo constituyen los estudios de regeneración que se están llevando a cabo en anfibios y en pollos. Es una observación muy antigua que los anfibios y reptiles son capaces de regenerar las extremidades mientras que las aves o los mamíferos no. Los estudios en marcha están permitiendo identificar los genes relacionados con el proceso regenerativo, varios de los cuales están también presentes en especies que no regeneran. Parece ser que la facultad de regenerar un órgano o no hacerlo depende no tanto de la presencia o ausencia de uno o varios genes como del mecanismo de regulación de genes comunes. Las especies que regeneran son capaces de activar estos genes después de un trauma físico, mientras que las que no regeneran no lo son. Una especulación bien fundada es que, cuando se conozca bien el proceso de regulación de estos genes, se podría intervenir en el control de su función con objeto de inducir artificialmente el proceso regenerativo en especies como la humana, que de forma natural no lo harían.

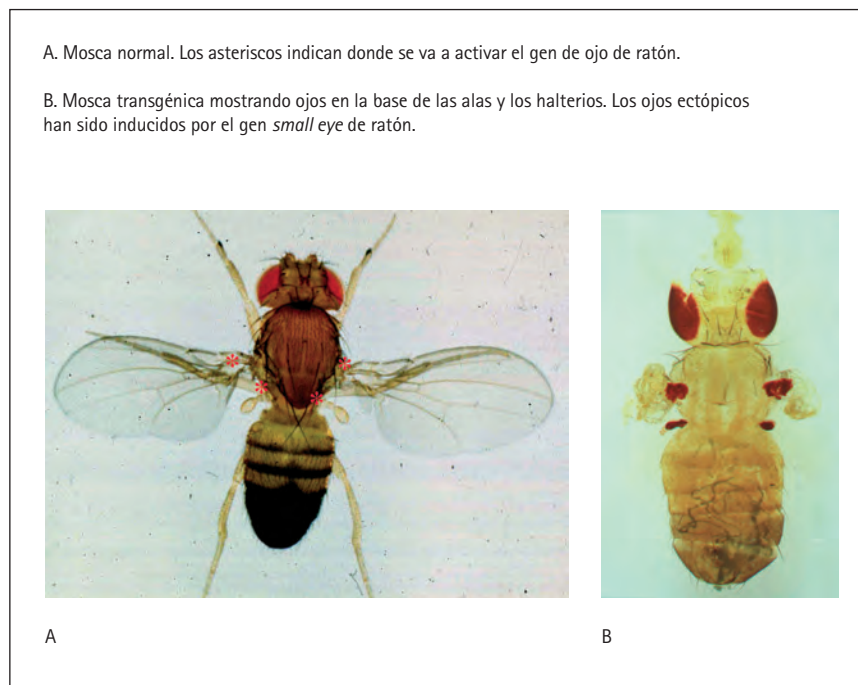


Figura 4. Conservación funcional de la programación genética del ojo.



Figura 5. Presentación oficial del Genoma Humano por el presidente Clinton (al centro de la fotografía). La finalización del Proyecto Genoma Humano fue posible gracias a la colaboración del consorcio público, liderado por Collins (a la derecha), y del privado, liderado por Venter (a la izquierda).

Los Proyectos Genoma

La exposición anterior es en sí misma una reivindicación de todo el fenómeno evolutivo, al exponer claramente la universalidad funcional de los fenómenos biológicos. Pero, además, las nuevas tecnologías moleculares nos han proporcionado una demostración más directa de esta universalidad. Durante los últimos años se ha terminado la secuenciación completa del ADN —los Proyectos Genoma— de muchas especies animales y vegetales, lo cual ha permitido comparar directamente el grado de similitud o diferencia de la información biológica entre especies diferentes.

En este contexto son particularmente relevantes los genomas de nemátodo *Caenorabditis elegans*, que contiene un ADN de 90 millones de pares de bases, el de la mosca *Drosophila* con 120 millones de pares de bases y el genoma humano con 3.300 millones de pares de bases. En el proyecto Genoma Humano (fig. 5) se utilizó el ADN de cinco personas —tres mujeres y dos hombres— de cuatro grupos étnicos diferentes —hispano, asiático, afro-americano y caucásiano—. Es interesante destacar que no se detectaron diferencias significativas entre ellos.

El gusano nemátodo <i>C. elegans</i> (1998)	90 x 106 bp	19.000 genes
La mosca del vinagre <i>Drosophila</i> (2000)	120 x 106 bp	14.000 genes
La especie humana (2001)	3.300 x 106 bp	40.000 genes

Las técnicas informáticas nos permiten comparar las secuencias de ADN de especies diferentes, es decir, su grado de similitud genética:

Los humanos tenemos 50% de identidad con el gusano *C. elegans*
60% de identidad con la mosca *Drosophila*

Estos proyectos han conseguido identificar todos los genes de cada especie, determinar su secuencia y acumular esta información en bases de datos, lo cual, junto con el desarrollo de instrumentos informáticos muy sofisticados y poderosos ordenadores, ha permitido comparar las secuencias significativas. La comparación ha producido muchos resultados de interés, pero uno particularmente importante (fig. 6) es el descubrimiento de que la especie humana comparte aproximadamente un 50% de los genes con el nemátodo *Caenorabditis elegans* y un 60% con la mosca *Drosophila*. Esta observación es un saludable recordatorio de nuestros orígenes biológicos, que compartimos con el resto de los animales. Naturalmente, esto se refleja en el ADN que es el trazo evolutivo común que nos une a todos.

Estudio de la enfermedad humana en organismos modelo

El alto grado de similitud genética en las especies mencionadas y, de hecho, en todo el Reino Animal no sólo valida el fenómeno evolutivo, sino que también tiene implicaciones poderosas en el estudio de la biología y patología humana. Al tener tantos genes en común con organismos como *Drosophila* hay muchos aspectos de la biología y de la enfermedad humana que se pueden estudiar en moscas sin las limitaciones experimentales y éticas que impone el material humano. La filosofía que subyace es que mucho del progreso en el conocimiento que consigamos en *Drosophila* será aplicable a nosotros mismos. Como vimos anteriormente, el estudio de los genes Hox de las moscas está arrojando información muy importante sobre la función de esos mismos genes en nuestra propia especie.

En lo que respecta a los procesos patológicos, las últimas estimaciones indican que el 74% de los genes relacionados con enfermedades humanas están presentes en *Drosophila*. Se trata, por tanto, de una fuente de información de enorme importancia para el conocimiento básico de la enfermedad humana. Actualmente, muchos laboratorios en todo el mundo están usando *Drosophila* como organismo para estudiar patologías como el cáncer, la enfermedad de Alzheimer, ataxias, etc. Un ejemplo de este enfoque lo constituyen los experimentos para inducir el síndrome molecular de la enfermedad de Alzheimer en *Drosophila*. La deposición de la proteína amiloide (A β) en las neuronas es una característica de la enfermedad. La forma patológica contiene 42 aminoácidos en vez de 40 y forma agregados llamados placas amiloides. La tecnología de *Drosophila* permite inducir la enfermedad en los ojos o el cerebro de la mosca y estudiar la evolución de la enfermedad. Se pueden producir cientos de individuos y probar gran cantidad de posibles remedios o compuestos que interfieren con el desarrollo de la enfermedad. Estos experimentos han permitido identificar una droga —Congo Red— que mitiga grandemente el efecto de la enfermedad en moscas. Aunque desgraciadamente esta droga es

Figura 6. Comparación de algunos genomas importantes.

tóxica para los humanos y no serviría para el tratamiento de la enfermedad, indica claramente el potencial de este tipo de tecnología. Experimentos de este tipo ya han identificado varias drogas dirigidas a tratar el cáncer y otros procesos degenerativos.

¿Se puede alterar la duración de la vida humana?

El altísimo grado de conservación en todo el Reino Animal de los fenómenos biológicos fundamentales permite especular sobre las posibilidades de manipular procesos considerados, hasta hace poco, inaccesibles a la intervención humana. Uno de los paradigmas fundamentales de la sociedad y cultura humana es la idea de que el envejecimiento y la muerte son procesos biológicos inevitables. Se supone que existe una programación interna que establece, dentro de un rango relativamente estrecho, la duración máxima de la vida de los individuos de cada especie.

Durante el siglo xx la duración media de la vida humana ha aumentado considerablemente, principalmente debido a la mejoría en el nivel de vida, las condiciones higiénicas y el progreso en medicina, pero se estima que la duración máxima está en torno a los 120-125 años. ¿Se podría sobrepasar este límite? Éste es un tema que ha suscitado gran atención en las revistas científicas internacionales (véase *Nature* 458, 1065-1071, 2008), debido fundamentalmente a que recientemente se han hecho descubrimientos que tienen que ver directamente con la programación genética de la duración de la vida.

El hecho fundamental es que se han identificado en el gusano nemátodo *Caenorhabditis elegans* y en la mosca *Drosophila* varios genes cuya función está directamente relacionada con el programa de envejecimiento de estas especies. Dada la facilidad de manipulación genética en estos organismos, se ha conseguido alargar de forma sustancial la vida en individuos de estas especies. En el caso del nemátodo se han conseguido gusanos que llegan a vivir entre seis y siete veces más de lo normal. Extrapolando a la especie humana, la vida media de la población sería de unos 350 años y habría individuos que sobrepasarían los 500.

Un aspecto importante de estos descubrimientos es que los genes de envejecimiento identificados en el gusano nemátodo y en *Drosophila* están también presentes en la especie humana. El mejor estudiado de estos genes, llamado DAF-16 en el gusano y FOXO en *Drosophila* y en humanos, está relacionado con la vía de la insulina y se han detectado formas variantes de FOXO que parecen ser particularmente frecuentes en individuos mayores de cien años. Mutaciones en la especie humana que afectan la

actividad de la vía de la insulina también se han detectado en individuos centenarios.

DAF-16/FOXO ha sido clonado y se han construido gusanos modificados genéticamente, en los que se han alterado los niveles funcionales de este gen y que se traducen en alteraciones que llegan a duplicar la duración de la vida de los gusanos. El hecho de que manipulando un solo gen se logre este resultado ilustra el potencial de estas técnicas.

Como se mencionaba anteriormente, este gen está presente en nuestra propia especie, lo que abre la posibilidad de que su manipulación pudiera utilizarse para modificar la duración de la vida de los seres humanos.

La futura evolución de la especie humana: el hombre tecnológico

Para terminar, me gustaría reflexionar brevemente sobre la propia evolución de la vida en el planeta y la de la especie humana. La vida en el planeta comenzó hace 2.000-3.000 millones de años. Los animales Bilateralía —los grupos animales existentes en la actualidad—, aparecieron hace 540 millones de años. Hace unos 100.000-200.000 años la selección darwiniana dio origen a la especie humana, como ha originado muchos millones de especies más, vivas o extintas. Sin embargo, el desarrollo intelectual/tecnológico de nuestra especie la ha hecho esencialmente inmune al proceso de selección natural, por lo tanto las reglas evolutivas normales nos afectan poco o nada en la actualidad.

La cultura humana comenzó hace unos 10.000 años y el desarrollo tecnológico hace doscientos años, mientras que la tecnología del ADN comenzó hace veinticinco. El progreso de esta tecnología ha sido rapidísimo y ha dado lugar a métodos muy poderosos de manipulación. Esto es, el vehículo de la evolución, el ADN, está siendo modificado directamente por la intervención humana. Estas metodologías, todavía muy crudas, se están usando en animales de experimentación, moscas, ratones, gusanos, etc., pero dada la gran similitud genética no está lejos el día en que se podrán aplicar a la especie humana. El potencial de estos métodos es enorme, sobre todo si se considera que comenzaron hace veinticinco años. Resulta imposible imaginar lo que serán capaces de realizar dentro de cincuenta años, por no hablar de 500 o 5.000. La especie humana tendrá la posibilidad de modificarse genéticamente a sí misma de forma controlada. Esta perspectiva ofrece enormes posibilidades de determinar su propio futuro biológico, su propia evolución. La tecnología del ADN ofrece un nuevo paradigma social; podrá cambiar completamente la propia esencia del ser humano.

la biomedicina en el cambio de siglo

JESÚS AVILA Y JOSÉ M. MATO

En el siglo anterior se produjo el paso de la biología descriptiva a la biología cuantitativa y de la medicina del diagnóstico a simple vista, a la medicina de las analíticas y protocolos. En ambos casos se buscaba una nueva metodología, más rigurosa y que diera datos reproducibles. Cuantificando y midiendo, se buscaba incrementar el carácter científico de la biología y de la medicina.

De acuerdo con estos criterios, basados en hacer correctas mediciones, se desarrolló la bioquímica y, posteriormente la biología molecular, utilizando, al principio, modelos simples para los análisis, y después modelos cada vez más complicados que permitieran, mediante extrapolación, conocer el funcionamiento del ser humano.

Así pues, la interpretación de una complejidad cada vez mayor fue realizándose partiendo del estudio de procesos simples, sin muchas variables, para después realizar el estudio de las interacciones entre estos procesos, lo que en la transición de los siglos **xx** y **xxi** se denominó biología de sistemas.

Novedades en la transición

En el tiempo que ha comprendido la transición entre los dos siglos, se logró descifrar la información –secuenciar– del genoma de diferentes organismos, entre ellos el del ser humano (Venter, Adams, Myers et al. 2001, 1304), habiéndose realizado una descripción, no muy rigurosa, de las proteínas que se expresan en esos organismos, tanto en lo referente a su cantidad como a la naturaleza de dichas proteínas.

Por otra parte, se conoció, a nivel básico, algo más sobre los elementos básicos de la vida. El Dogma Central de la biología molecular indica que el genoma (DNA), se transcribe en RNA y una forma de éste, el RNA mensajero, se traduce dando lugar a la síntesis de proteínas. En estos últimos años, aparte del RNA mensajero, RNA de transferencia y RNA ribosómico, se ha descrito un nuevo tipo de RNA, el RNA de interferencia que actúa como regulador de la expresión génica (Fire, Xu, Montgomery et al. 1998, 806).

Sin embargo, la segunda parte del código genético, la clave de cómo se pliegan las proteínas sigue sin conocerse.

Este hecho es importante pues existen varias enfermedades, las denominadas proteinopatías, cuya base está en un plegamiento defectuoso, y no funcional, de una proteína, que, en muchos casos da lugar a la agregación aberrante de dichas proteínas.

Niveles de estudio

Se ha estudiado no sólo el nivel molecular, sino el nivel celular. Mientras que los procesos moleculares básicos son comunes no sólo entre las células de un mismo organismo, sino de organismo a organismo, la gran variedad celular en un organismo superior, como un mamífero, ha dado lugar a estudios específicos y especializados para los diferentes tipos celulares. Hay células que proliferan siguiendo un ciclo determinado que cuando funciona aberrantemente puede dar lugar a la aparición de tumores.

Hay células que sobreviven mucho tiempo sin dividirse y en estado diferenciado, como las neuronas. Hay células que interactúan íntimamente entre ellas, como las que forman el tejido epitelial, mientras las del tejido conectivo se rodean de una matriz extracelular. Pero, en general, el proceso de diferenciación empieza con unas células precursoras que tras diversas modificaciones dan lugar a las células maduras. Estas células precursoras, proliferantes, derivan a su vez de unas células más primitivas denominadas troncales o células madre. En este cambio de siglo se ha realizado un importante avance en el conocimiento de estas células madre. En este sentido se han desarrollado protocolos para diferenciar células madre de origen embrionario humano en diferentes tipos celulares: tejido muscular, tejido nervioso, etc. (Thomson, Itskovitz-Eldor, Shapiro et al. 1998, 1145) sugiriéndose que este proceso puede ser utilizado como terapia de regeneración celular. Por otra parte, hay largas discusiones sobre la posibilidad de que células madre de algún tejido adulto pudieran ser pluripotentes, es decir, pudieran diferenciarse a diferentes tipos de tejidos, discusiones que siguen hoy en día. Más recientemente se ha descrito un descubrimiento que podría ser interesante. Este descubrimiento, reprogramación, ha consistido en revertir la diferenciación de células maduras, de tejidos específicos, convirtiéndolas en células con características de células madre embrionarias. La expresión los factores de transcripción Oct4, Sox2, Klf4 y cMyc convierte, reprograma, fibroblastos diferenciados en células con características similares a las de las células madre (Takahashi y Yamanaka 2006, 663). Esta reprogramación podría darse de un modo natural en la transición entre células epiteliales y mesenquimales que tienen lugar durante el desarrollo (Mani, Guo, Liao et al. 2008, 704). Posteriormente, estas células con características de células troncales pueden diferenciarse nuevamente en tipos celulares diferentes del inicial. En la rata, neuronas obtenidas de fibroblastos reprogramados se pueden transplantar en un ratón con síntomas de la enfermedad de Parkinson, produciendo en éstos una recuperación funcional (Wernig, Zhao, Pruszak et al. 2008, 5856).

Nuevas terapias

A partir de los dos descubrimientos anteriormente indicados relativos al RNA de interferencia, RNAi, y a la caracterización de las células madre, se han establecido métodos de terapia molecular, basados en el uso de los RNAi, para impedir la expresión de una proteína que pueda ser tóxica para el organismo; o métodos de terapia celular mediante el uso de células precursoras en procesos regenerativos. Una fuente alternativa de células troncales es la sangre del cordón umbilical (Kurtzberg, Laughlin, Graham et al. 1996, 157). Estas células pueden ser utilizadas para terapia celular en enfermedades como leucemias o hemoglobinopatías, habiéndose establecido bancos para almacenar estas células.

La regeneración de órganos y tejidos varía drásticamente con la naturaleza de las células que forman parte de estos tejidos. Es conocido que, tras una hepatotomía, el hígado regenera hasta su tamaño inicial, como también se conoce que puede haber un recambio de las células de la sangre, de la piel o de los huesos, de un modo natural. Artificialmente, en el caso de las células de la piel, se ha conseguido obtener piel cultivada con propósitos regenerativos (O'Connor, Mulliken, Banks-Schlegel et al. 1981, 75), tras realizar cultivos celulares de queratinocitos y fibroblastos. También el implante de condriocitos ha permitido la regeneración del cartilago (Brittberg, Lindahl, Nilsson et al. 1994, 889). Más difícil es la regeneración en otros órganos, como el cerebro.

El desarrollo de áreas como la biología molecular, la biología celular o la biotecnología ha permitido diagnosticar mejor diferentes enfermedades y, en algunos casos, proponer tratamientos personalizados para los pacientes. Son claras las diferencias que pueden observarse entre los seres humanos; diferentes géneros, razas, colores de pelo, estatura, volumen corporal, etc. Esto se debe, en parte, a diferencias en sus genomas, DNA, y recientemente se ha observado que aunque existen los denominados polimorfismos de un solo nucleótido, que dan lugar a la existencia de pequeñas diferencias de un mismo gen en diferentes individuos, parece ser que las mayores diferencias se deben a la existencia de deleciones y/o inversiones en el material genético de un individuo comparado con el material genético de otro individuo. Esta observación se consideró por la revista *Science* como uno de los hitos científicos del pasado año (Kennedy 2007, 1833).

Difusión de los resultados

Es importante resaltar la incidencia mediática en la mayor o menor visualización de descubrimientos. Como se ha indicado, al final de cada año la revista *Science* resalta sus descubrimientos favoritos. También *Nature*, a través de la revista *Nature Methods*, indica los métodos del año. Este año el principal método fue una nueva tecnología para secuenciar, más rápido y barato, genomas como el de diferentes seres humanos (Wold y Myers 2008, 19). Se ha sugerido que con esta técnica se ha secuenciado el genoma de los científicos Watson y Venter.

También la revista *Cell* hace propaganda de los descubrimientos que publica en sus páginas. Hace pocos años, coincidiendo con el 30 aniversario de la revista, se recopiló en un número especial (*Cell 30th Anniversary Supplement*, January 23, 2004), los trabajos más relevantes previamente publicados en *Cell*, entre ellos varios por los que sus autores fueron galardonados con el Premio Nobel. Un ejemplo es el referente a los mecanismos de degradación de las proteínas del citoplasma a través de una maquinaria conocida como proteasoma (Rader y Daugherty 2008, 904). Es posible que algunos de los Premios Nobel futuros correspondan a algún autor con alguna contribución en dicho número.

Las enfermedades más prevalentes

Por otra parte, el conocimiento molecular de diversos tipos celulares encontrados en diferentes tejidos puede facilitar el mejor conocimiento de enfermedades. Desde un punto de vista muy general, se puede considerar que en un organismo existen cuatro tipos mayoritarios de tejidos: el muscular, el nervioso, el epitelial y el conectivo. El primero está muy relacionado con problemas cardiovasculares, dado el carácter muscular del corazón; el segundo con problemas neurodegenerativos; el tercero con procesos infecciosos y parte de este último y el cuarto con la mayor formación de tumores. La solución de estos cuatro tipos de problemas; cardiovasculares, neurodegenerativos, oncológicos e infecciosos son los retos fundamentales de la medicina del siglo XXI. En este capítulo nos ocuparemos, principalmente, de problemas relacionados con defectos metabólicos o con procesos neurodegenerativos.

En aspectos relacionados con enfermedades cardiovasculares referimos al lector a un número específico de *Nature*, volumen 451, issue 7181, casi monográfico sobre el tema, en donde se comentan posibles nuevas terapias para arterioesclerosis (Rader y Daugherty 2008, 904), posibles tratamientos para la trombosis (Mackman 2008, 914), o el uso de células madre para enfermedades cardíacas (Sengers y Lee 2008, 937).

Respecto a aspectos relacionados con cáncer referimos al lector al número de *Nature* 441, 7092; que incluye el suplemento «Insight: signalling in cancer», donde hay interesantes artículos como el de Benson J. D. et al. sobre la validación de nuevas dianas para ensayo de compuestos antitumorales.

Antes de acabar esta parte introductoria, queremos comentar brevemente el uso de modelos animales para el estudio de enfermedades y algunas mejoras en el diagnóstico médico, desde un punto de vista general.

Modelos animales

En muchos casos se han intentado reproducir algunos aspectos patológicos de diversas enfermedades utilizando modelos animales que minifiquen todas o algunas características de la enfermedad. Estos modelos van desde el gusano al ratón, pasando por la mosca. Fundamentalmente, los estudios en los que se ha utilizado ratones como modelos, han servido para usar estos modelos como dianas terapéuticas para probar fármacos, como primer paso para su utilización posterior en la clínica. El uso de modelos de ratón ha sido reconocido con la concesión del Premio Nobel de Medicina en 2007.

La mejora en el diagnóstico médico

Sin embargo, aparte de los modelos, es muy importante conocer qué pasa específicamente en el cuerpo de un ser humano enfermo. En este sentido se ha aumentado en gran proporción el conocimiento en el diagnóstico médico. Éste se puede obtener tras analizar los componentes

de fluidos como sangre, plasma, orina o líquido cefalorraquídeo; pero quizás el mayor avance puede haber sido el desarrollo de nuevas técnicas de imagen, como por ejemplo la resonancia magnética de imagen funcional, que está aportando una excelente información de zonas difíciles de analizar en un ser humano, como es el cerebro (Kerr y Denk 2008, 195).

Como hemos indicado anteriormente, vamos a enfocar fundamentalmente nuestro trabajo a comentar, más específicamente, algunos aspectos sobre procesos neurodegenerativos o problemas de metabolismo.

La enfermedad de Alzheimer

Como ejemplo de proceso neurodegenerativo comentamos la enfermedad de Alzheimer, una demencia senil que se caracteriza, inicialmente, por una pérdida de memoria a la que sucede, con el progreso de la enfermedad, una gran desorientación, demencia, del paciente (Alzheimer 1907, 146).

El envejecimiento es el mayor riesgo para padecer la enfermedad de Alzheimer. Su prevalencia se duplica cada cinco años, a partir de los 65 años de edad. Casi un 2,5% de personas de 65 años la padecen, mientras la incidencia es entre 40 y 50% en personas mayores de 85 años.

Alrededor de la enfermedad surgen diferentes problemas que se añaden al esencial, la pérdida de memoria, el entendimiento y la voluntad del paciente. Como consecuencia de este deterioro se requiere el cuidado del paciente por otras personas, generalmente los familiares más próximos que, posiblemente, son los más perjudicados por la enfermedad. Éstos son los problemas fundamentales, a los que se les añaden otros de ámbito social y económico.

Datos de la Organización Mundial de la Salud, de 2001 (Vas 2001), estimaban que la enfermedad de Alzheimer afectaba a 18 millones de personas en el mundo. Datos aproximados del año 2006, subían a 26 millones de personas el número de enfermos a nivel mundial. Se ha estimado que esta cifra podría multiplicarse por tres en el año 2050, dada la mayor longevidad del actual ser humano y al hecho de que el mayor riesgo para padecer la enfermedad es el envejecimiento, según lo publicado en la página web de la Alzheimer's Association con el título *Alzheimer's Disease Facts and Figures* en 2007 (http://www.alz.org/alzheimers_disease_facts_figures.asp). En este sentido se ha sugerido también que en personas centenarias la probabilidad de padecer demencia senil pudiera ser mayor del 50%. Es decir, la neurodegeneración podría considerarse como un proceso normal, que ocurriría a edades muy avanzadas.

Desde un punto de vista económico, en Estados Unidos se ha calculado que el cuidado de 4,5 millones de pacientes puede tener un coste mínimo de cien mil millones de dólares (Glennier y Wong 1984, 1131, Masters, Simms, Weinman et al. 1985, 4245). Por otra parte, datos de 2006, indican un gasto en fármacos paliativos, los únicos que existen actualmente al carecer de curativos, de 4.600 millones de dólares.

Así pues, es una enfermedad crónica, devastadora y con un gran coste humano, social y económico. Una enfermedad que, además, va aumentando su prevalencia, debido al incremento de la edad media de la población.

Desde el año 1907, en el que A. Alzheimer descubrió el primer caso de la enfermedad, hasta la década de lo ochenta del siglo xx, lo que se conocía de la enfermedad de Alzheimer fundamentalmente derivaba del comportamiento anómalo de los pacientes y de los estudios histopatológicos de sus autopsias. La autopsia del cerebro de un paciente de Alzheimer indica una abundancia poco usual de dos tipos de estructuras aberrantes, las placas seniles y los ovillos neurofibrilares. Además, las autopsias de los pacientes muestran una gran mortandad de neuronas, principalmente en el hipocampo, la zona límbica y en la corteza cerebral.

A partir de los años ochenta del siglo pasado, se ha empezado a estudiar qué sucede en los cerebros de los pacientes de Alzheimer que conduce a la formación de placas seniles e, independientemente, de los ovillos neurofibrilares. Se descubrió que el componente mayoritario de las placas seniles era un péptido (Glenner y Wong 1984, 1131), del que se obtuvo su secuencia de aminoácidos y, dado que las placas seniles tienen forma de agregados de amiloide, al péptido se le denominó como péptido beta amiloide. Posteriormente, se encontró que este péptido era un fragmento de una proteína (Masters, Simms, Weinman et al. 1985, 4245), a la que se denominó proteína precursora del amiloide, APP. Esta proteína está oblicuamente presente en distintos tipos celulares, pero el péptido beta amiloide sólo produce agregados aberrantes en el sistema nervioso. Estudiando las características del péptido amiloide se llegó a la conclusión de que los cortes que tenían lugar en la proteína APP para producirlo, no eran los usuales que tenían lugar en esta proteína en situaciones no patológicas. En esas situaciones, APP, cuya localización celular es la membrana plasmática, seguía unida a la membrana o se desligaba mediante el corte de una proteasa, denominada secretasa α , pues tras el corte, el fragmento mayoritario se secreta al medio extracelular. Sin embargo, en condiciones patológicas (Price, Tanzi, Borchelt et al. 1998, 461), en vez de cortarse la proteína APP por la secretasa α , se corta por otra proteasa, denominada secretasa β , y posteriormente por otra proteasa, secretasa γ . El corte de APP por las secretasas α y β da lugar al péptido amiloide, que, cuando se acumula, da lugar a diferentes tipos de agregados siendo los de mayor tamaño las placas seniles. Mientras el corte debido a la beta secretasa es específico, y se produce entre dos aminoácidos determinados, el corte de APP por la secretasa γ es impreciso, dentro de una región determinada de la proteína. Dada esta imprecisión, se obtienen péptidos beta amiloide de diferente tamaño, siendo los más usuales los péptidos que contienen 40, $A\beta_{40}$, y 42, $A\beta_{42}$, residuos. El segundo posee una mayor capacidad de agregación que el primero. Así pues, $A\beta_{42}$ se le considera

el péptido amiloide con mayor potencial tóxico. Una vez formados los péptidos beta amiloide, éstos pueden ser degradados por proteasas como la neprelisina o la enzima que degrada la insulina, IDE, que, al degradar a estos péptidos beta amiloide, previene su agregación aberrante, que puede ser tóxica. Esta toxicidad ha sido observada cuando el péptido se añade a un cultivo de células neuronales. La toxicidad del péptido beta amiloide agregado puede ser debida a que facilite la entrada de calcio al citoplasma celular, y/o a actuar como antagonista en algunas vías de señalización en las neuronas. Además, se ha sugerido una acción sobre las células de microglia, que facilitaría la secreción, por estas células de glia, de citoquinas y otros factores, que provocarían un proceso inflamatorio que podría concluir con muerte neuronal.

Adicionalmente, se ha sugerido que para que el efecto tóxico del péptido amiloide tenga lugar en neuronas, se requiere, en dichas neuronas, la presencia de la proteína tau (véase abajo).

El componente mayoritario de los ovillos neurofibrilares es una proteína asociada a los microtúbulos, denominada tau (Grundke-Iqbal, Iqbal, Tung et al. 1986, 4913). Esta proteína, modificada por hiperfosforilación está presente en los denominados filamentos apareados helicoidales, PHF, cuyos agregados son los ovillos neurofibrilares. Utilizando la proteína tau aislada, se pudo comprobar que ella sola era suficiente para producir agregados similares a los PHF (Montejo de Garcini, Serrano y Avila 1986, 790). Por otra parte se ha descrito que diferentes proteína quinasas pueden modificar, fosforilar, a la proteína tau, siendo la que modifica más residuos de la proteína tau la quinasa conocida como GSK3 (Avila, Lucas, Pérez et al. 2004, 361). Se sugiere que tanto la proteína tau fosforilada como los agregados que se forman a partir de la proteína tau, pueden ser tóxicos para las células en donde estén presentes.

Más recientemente, se ha sugerido que tras la muerte neuronal, tau intracelular pasa al medio extracelular y que este tau extracelular puede ser tóxico para las neuronas cercanas, contribuyendo, de esta manera, a la propagación de la patología (Gómez-Ramos, Díaz-Hernández, Rubio et al. 2008, 673).

La enfermedad de Alzheimer se ha dividido en dos tipos, la de origen genético o enfermedad de Alzheimer familiar, y la de origen esporádico. La primera tiene muy poca incidencia, posiblemente menos del 1% del total de los casos de la enfermedad son de origen familiar, por lo que el tipo más usual de la enfermedad es el esporádico.

Sin embargo, el conocimiento del mecanismo de la enfermedad de Alzheimer familiar ha dado importantes pistas sobre la enfermedad en general. Éste es un hecho importante ya que si se conoce el proceso de cómo se forman las placas y los ovillos se podrán, posiblemente, diseñar terapias para combatir la enfermedad.

Los casos familiares de la enfermedad de Alzheimer se deben a mutaciones en tres genes diferentes que codifi-

can tres proteínas: APP, la proteína precursora del péptido amiloide, presenilina 1, PS-1, y presenilina 2, PS-2. Las mutaciones de la proteína APP, que inducen el desarrollo de la enfermedad, facilitan su corte por las secretasas beta y gamma, y dificultan el corte por la secretasa α . En todos estos casos, se facilita la formación del péptido beta amiloide (Price, Tanzi, Borchelt et al. 1998, 461).

Por otra parte, se analizó la naturaleza de la γ secretasa, observándose que está formada por un complejo de cuatro proteínas, una que posee la capacidad catalítica proteasa, y tres proteínas adicionales que regulan la actividad catalítica. La proteína con la actividad catalítica es PS-1 o PS-2, que pueden intercambiarse, y, mayoritariamente, las mutaciones en PS-1 y en PS-2, que inducen la aparición de la enfermedad, dan lugar a una ganancia en la función proteolítica de estas proteínas. Es decir, las mutaciones que se encuentran en PS-1 o PS-2, en pacientes con la enfermedad de Alzheimer de origen familiar, favorecen la aparición del péptido beta amiloide (Price, Tanzi, Borchelt et al. 1998, 461).

Como en casi todos los casos las mutaciones en APP, PS-1 o PS-2 dan lugar a un aumento en la producción del péptido beta amiloide, se sugirió, en la «hipótesis de la cascada del amiloide» (Hardy y Selkoe 2002, 353), que el primer paso para el desarrollo de la aparición de la enfermedad era la presencia, en una determinada cantidad, del péptido amiloide, que, tras agregar, podría originar los procesos patológicos posteriores, pudiéndose encontrar entre estos procesos posteriores la hiperfosforilación y agregación de la proteína tau. Sin embargo, el análisis a nivel de anatomía patológica del desarrollo de la enfermedad no apoyó esta hipótesis pues, aparentemente, la patología relacionada con la proteína tau y no la relacionada con la proteína amiloide parece que se correlacionan más con el proceso patológico de la enfermedad (Braak y Braak, 1991, 239, Arriagada, Growdon, Hedley-Whyte et al. 1992, 631). Por ello, se analizó si el resultado de las mutaciones en APP, PS-1 y PS-2 pudiera converger en la modificación de alguna otra proteína. Una posible proteína es la proteína quinasa GSK3, pues mutaciones en APP, que dan lugar a la aparición del péptido beta amiloide, facilitan la activación de la actividad quinasa de GSK3, ya que el péptido beta amiloide actúa como antagonista de vías de señalización, insulina o wnt, que dan lugar a una inactivación de la GSK3 (Avila, Lucas, Pérez et al. 2004, 361). Por otra parte, mutaciones en PS-1 o PS-2 que dan lugar a un aumento en la cantidad del péptido amiloide pueden tener iguales consecuencias a las indicadas para APP mutada, mientras aquellas mutaciones en PS1/PS2 que no dan lugar a un aumento en la producción del péptido beta amiloide pueden, por otras vías, aumentar la actividad de GSK3 (Baki, Shioi, Wen et al. 2004, 2586).

Dada la confluencia de las mutaciones de APP, PS-1 y PS-2 en un efecto de activación de GSK3, se desarrolló un modelo de ratón transgénico que sobreexpresaba la quinasa en aquellas zonas, hipocampo y corteza, más afectadas en

la enfermedad de Alzheimer (Lucas, Hernández, Gómez-Ramos et al. 2001, 27). Este ratón reproducía algunos aspectos de la patología de tau y, además, mostraba déficits de memoria (Hernández, Borrell, Guaza et al. 2002, 1529), por lo que se ha utilizado como diana para probar fármacos que pudieran inhibir la actividad quinasa y que, como consecuencia, pudieran reparar el déficit cognitivo. En estos ratones modificados genéticamente, la lesión más evidente es la degeneración del giro dentado (Engel, Lucas, Gómez-Ramos et al. 2006, 1258), algo que sucede también en los enfermos de Alzheimer, y que puede ser la responsable de la pérdida de memoria observada, tanto en el modelo animal como en los pacientes de Alzheimer.

Clínicamente los enfermos tienen, inicialmente, una pérdida creciente de memoria, deterioro cognitivo leve, que se ha relacionado con lesiones en la región hipocámpal, donde se localiza el giro dentado. Posteriormente, la patología se extiende por la zona límbica del lóbulo temporal, y más tarde a la corteza frontal, dando lugar a problemas de consolidación de la memoria, de comportamiento y de lenguaje. Después se puede observar muerte neuronal en la corteza parietal, lo que puede dar lugar a problemas visual-espaciales o de desorientación, por ejemplo, en el uso de utensilios, o a problemas de incapacidad para tomar decisiones, en donde junto con la corteza parietal también interviene la corteza frontal. Todos estos problemas relacionados con desorientación dan lugar a lo que, clínicamente, se conoce como demencia. Así pues, la enfermedad podría dividirse, de una forma muy general, en dos grandes estadios; uno inicial relacionado con la pérdida de memoria y otro posterior relacionado con la aparición de la demencia. Los dos problemas son de suma importancia, pero el segundo requiere una mayor atención por parte de los cuidadores de los pacientes.

En la transición entre el siglo xx y el xxi se ha producido un gran avance en el conocimiento, a nivel básico, de la enfermedad. Sin embargo, todavía no hay una buena aplicación terapéutica de estos conocimientos que permita combatir la enfermedad.

Como posibles terapias, hasta el presente se han utilizado fármacos paliativos más que curativos o modificadores. Estos fármacos retrasan tímidamente el desarrollo de la enfermedad que, desgraciadamente, finalmente progresa. Los fármacos, aprobados por la Agencia Norteamericana del Medicamento, FDA, son: tacrina, donepezilo, galantamina, rivastigmina y memantina. Sólo la venta anual del donepezilo ronda los 1.000 millones de dólares, mientras que las ventas de memantina, el más recientemente incorporado al mercado, son de alrededor de los 500 millones de dólares (Mount y Downton 2006, 780, Stephen Salloway 2008, 65). Los cuatro primeros son inhibidores de la enzima acetilcolinesterasa (Stephen Salloway 2008, 65). Esta enzima degrada el neurotransmisor acetilcolina, requerido para la perfecta transmisión neuronal. En la enfermedad de Alzheimer hay un daño prefe-

rencial de las neuronas colinérgicas, aquellas que utilizan acetilcolina como neurotransmisor. Con objeto de intentar mantener altos los niveles del neurotransmisor es por lo que se emplean estos fármacos, aunque el primero, tacrina, se cayó de la lista debido a su toxicidad. La memantina es un antagonista del receptor de otro neurotransmisor, el glutamato. Se ha observado que en neuronas de personas de la tercera edad, hay una excesiva activación de un tipo de receptores del glutamato, los receptores NMDA, activación que puede ser tóxica, dañando a la neurona. Con objeto de proteger a las neuronas de estas personas mayores se les administra memantina, que es un antagonista de los receptores NMDA (Parsons, Danysz y Quack 1999, 735) (Reisberg, Doody, Stoffler et al. 2003, 1333).

Éstos son los fármacos actuales. A continuación comentaremos brevemente algunos métodos de diagnóstico y cuáles son los posibles fármacos futuros.

Como biomarcadores de la enfermedad se pueden determinar los niveles de la proteína tau, con diferentes niveles de fosforilación, y del péptido beta amiloide, en el líquido cefalorraquídeo. Más recientemente, se han determinado los niveles de hasta 18 componentes del plasma como posibles indicadores de la enfermedad (Ray, Britschgi, Herbert et al. 2007, 1359), pero quizás los métodos diagnósticos que han recibido más atención han sido los relacionados con las técnicas de imagen como el PET (Blennow y Zetterberg 2006, 753) y la resonancia magnética funcional (Logothetis 2008, 869). Con estas técnicas se puede seguir el agrandamiento de los ventrículos, tras la muerte neuronal, que aqueja a los enfermos de Alzheimer. De las dos técnicas indicadas, la resonancia magnética funcional parece tener algunas ventajas. Esta técnica mide cambios hemodinámicos en diferentes puntos del cerebro y tiene las ventajas de no ser invasiva y de tener una buena resolución espacio-temporal, que puede mostrar un resultado que se correlaciona con una actividad específica realizada por un individuo (Logothetis 2008, 869).

Los fármacos que están ya, o muy próximos a estar, en ensayos clínicos y que pueden ser modificadores (Stephen Salloway 2008, 65) y no paliativos de la enfermedad, es decir los posibles fármacos del futuro para esta enfermedad, tienen un mecanismo basado en las diferentes observaciones realizadas, a nivel básico, sobre la patología de la enfermedad. Así pues se están desarrollando fármacos que reduzcan los niveles del péptido beta amiloide. De este modo se están desarrollando anticuerpos específicos, vacunas, contra el péptido amiloide –el producto se denomina Bapineuzumab–. Por otra parte, se están desarrollando inhibidores contra la beta y gamma secretasa; algunos de ellos moduladores de la γ secretasa, buscando reducir la relación $A\beta_{1-42}/A\beta_{1-40}$; y algunos muestran además un posible efecto antiinflamatorio, flurizan y tarenflumol. Sin embargo, noticias recientes muestran datos negativos para el flurizan. Hay otros compuestos que previenen la agregación del beta amiloide, clioquinol, o extractos de

polifenoles que también pueden prevenir la oligomerización del péptido beta amiloide (Wang, Ho, Zhao et al. 2008, 6388) (Stephen Salloway 2008, 65); o que pueden mantener altos los niveles de aquellas enzimas, como la enzima que degrada insulina, IDE, que pueden degradar al péptido amiloide. Entre estos compuestos está un agonista del PPAR γ , rosiglitazone (Pedersen, Mcmillan, Kulstad et al. 2006, 265). Adicionalmente, se han buscado inhibidores que no se unan directamente a la γ secretasa, sino a su sustrato, impidiendo el corte proteolítico de la enzima en dicho sustrato, pero no en otros (Kukar, Ladd, Bann et al. 2008, 925).

Por otra parte, se ha mostrado que el efecto tóxico de la excesiva actuación de los receptores NMDA puede incrementar los niveles del péptido beta amiloide (Harkany, Abraham, Timmerman et al. 2000, 2735) y de la proteína tau (Amadoro, Ciotti, Costanzi et al. 2006, 2892), por lo que aparte de la memantina se están buscando otros antagonistas de los receptores NMDA, siendo el dimebon uno de ellos. En otro tipo de estudios, no se han conseguido grandes logros en la búsqueda de antioxidantes que pudieran actuar como neuroprotectores.

Por último, referente a la patología relacionada con la proteína de tau, se están buscando inhibidores específicos de aquellas quinasas que, como GSK3, mayoritariamente modifican a dicha proteína tau. Más recientemente, se ha descrito que el azul de metileno, un antiagregante de la proteína tau, podría tener un efecto terapéutico.

Éstos son algunos ejemplos que pueden ser indicativos del tremendo esfuerzo que se está realizando para prevenir esta terrible enfermedad.

La enfermedad del hígado graso no-alcohólico

La enfermedad del hígado graso no-alcohólico –abreviado NAFLD, del inglés *non-alcoholic fatty liver disease*– es la causa más frecuente de enfermedad hepática en los países occidentalizados y, consecuentemente, un problema de salud mundial. Afecta a alrededor de 20 millones de personas en Estados Unidos, y a un número similar en la Unión Europea (Adams y Lindor 2007; Ahmed y Byrne 2007). NAFLD es frecuente en pacientes con obesidad, diabetes, hiperlipidemia e hipertensión (Abdelmalek y Diehl 2007). El incremento de la obesidad en los países occidentalizados justifica el creciente interés en el estudio de NAFLD. Aproximadamente el 50% de los individuos obesos tienen NAFLD (Angulo 2007). NAFLD es un término clínico-patológico que se emplea para describir un amplio rango de situaciones que van desde la simple acumulación de grasa en el hígado –esteatosis no-alcohólica– a la esteatohepatitis no-alcohólica –NASH, acumulación de grasa con inflamación, necrosis y fibrosis–. NAFLD es en general una enfermedad asintomática, aunque en una minoría de los pacientes con NAFLD la enfermedad progresa hasta desarrollar cirrosis, fallo hepático y carcinoma hepatocelular –HCC–. Aproximadamente el 10% de los pacientes con

NAFLD desarrollan NASH; y de éstos, entre el 10 y el 20% desarrollan cirrosis y HCC. El consumo excesivo de alcohol también produce hígado graso y, al igual que ocurre en NAFLD, también en este caso la enfermedad hepática puede progresar hasta la aparición de esteatohepatitis, cirrosis y HCC. No obstante lo anterior, es importante hacer hincapié en que NAFLD y la enfermedad hepática inducida por el consumo excesivo de alcohol son dos enfermedades distintas. NASH es también diferente de otras formas de hepatitis causadas por la infección con diversos virus, como el virus de la hepatitis B o C. Mientras que el diagnóstico clínico de NAFLD está basado en la elevación de las transaminasas en sangre, en el índice de masa corporal –IMC, se calcula dividiendo el peso de un individuo en kilogramos por la altura en metros al cuadrado; los valores entre 18,5 y 25 se consideran normales–, en la evidencia mediante ecografía o resonancia magnética de la acumulación hepática de grasa, y en la presencia de otros factores como obesidad, diabetes, hipertensión e hiperlipidemia; para confirmar la presencia de NASH y el grado de fibrosis y necrosis es necesario realizar una biopsia hepática.

A pesar de ser NAFLD un problema de salud mundial, no se sabe por qué en algunos individuos la enfermedad progresa a NASH y en otros no. Existen varias hipótesis. La principal tiene que ver con que son necesarios dos *hits*: el primer *hit* es la acumulación de grasa en el hígado, pero la naturaleza del segundo *hit* es una incógnita, aunque los estudios realizados en los últimos años empleando modelos animales genéticamente modificados han proporcionado nuevas claves sobre el segundo *hit*. Tampoco se sabe por qué en algunos pacientes NASH progresa a cirrosis y HCC, y en otros no.

Aproximadamente el 25% de la población adulta estadounidense es obesa, y un 55% adicional tiene sobrepeso (Sturm 2002). Aunque estos números son algo mejores en la población adulta europea, la obesidad en la Unión Europea es también un problema de salud pública. La obesidad está causada, fundamentalmente, por un exceso en la ingesta de calorías relativo al consumo de energía. A nivel individual, la obesidad puede ser más compleja e implicar factores genéticos que regulan el metabolismo y almacenamiento de los lípidos, el control por el cerebro de los hábitos de alimentación y de ejercicio, u otros factores desconocidos. Es cierto que en algunos individuos hay factores genéticos que influyen de manera determinante su disposición a acumular grasa o que favorecen la sensación de hambre frente a la de saciedad. Para estos individuos controlar el peso es muy difícil, pero para la mayoría de la población los factores genéticos son menos determinantes y pueden ser compensados, más fácilmente, mediante cambios en los hábitos de alimentación. Cualquier individuo obeso perderá peso si es forzado a mantener una dieta baja en calorías en combinación con la realización de ejercicio. Sabemos que esta afirmación es cierta, porque los pacientes que son sometidos a un *by-pass* gástrico

–mediante esta técnica quirúrgica el tamaño del estómago se reduce drásticamente desde alrededor de 1 litro hasta tan sólo 30-60 mililitros– reducen marcadamente la cantidad de grasa del tejido adiposo.

Para la mayoría de los individuos, el tejido adiposo no es más que una masa inerte de grasa pero, desde mediados de los años noventa, sabemos que es un tejido biológicamente muy activo. En 1994, Friedman y colaboradores identificaron la hormona leptina y descubrieron que su deficiencia era la causa de la obesidad extrema observada en un ratón mutante denominado *obeso* (Zhang y cols 1994). Estos ratones son enormes. Mientras que un ratón normal pesa alrededor de 30 gramos, los ratones *obesos* llegan a pesar 90. Estos ratones tienen elevados los niveles de lípidos en sangre y desarrollan hígado graso. La leptina se sintetiza en el tejido adiposo. En los animales normales, la cantidad de leptina en sangre es proporcional a la cantidad de tejido adiposo. Mediante este mecanismo, el tejido adiposo se comunica con el cerebro para informarle de que la ingesta de alimentos ha sido suficiente. Los animales *obesos* tienen una mutación en el gen de la leptina, de manera que sintetizan una hormona que no es biológicamente activa y, por consiguiente, el cerebro de estos animales no recibe la señal adecuada para dejar de comer. Desafortunadamente, en la mayoría de los individuos obesos la concentración de leptina se encuentra anormalmente elevada, no disminuida. Pero los pocos pacientes obesos que tienen una deficiencia genética de leptina responden bien al tratamiento con esta hormona reduciendo la acumulación de grasa en el cuerpo.

Aunque el descubrimiento de la leptina no condujo a la curación de la obesidad, sí cambió para siempre la manera de pensar acerca de la fisiopatología de la obesidad. Desde el descubrimiento de la leptina, se han descubierto otras hormonas y citoquinas –proteínas cuya actividad principal es el control de la inflamación– que tienen su origen en el tejido adiposo y que regulan el apetito y/o el metabolismo de los lípidos, como la adiponectina –una hormona sintetizada por el tejido adiposo que favorece la oxidación de los ácidos grasos y el metabolismo de la glucosa y cuyos niveles en sangre son inversamente proporcionales al IMC–, resistina –una hormona sintetizada por el tejido adiposo relacionada con la inflamación y la diabetes– y el factor de necrosis tumoral- α –TNF- α , una citoquina involucrada en la regulación de la muerte, diferenciación y proliferación celular, que juega un importante papel en la etiología de diversas enfermedades, incluida la diabetes–. Es decir, el tejido adiposo no funciona exclusivamente como un depósito de grasa, sino que desempeña también un papel fundamental en el control del apetito y del metabolismo de los lípidos e hidratos de carbono.

Las señales que impulsan a comer deben estar en equilibrio con el centro de control del apetito del cerebro, de manera que éste pone en marcha el deseo de comer de nuevo cuando se produce un balance energético negativo.

Los estudios llevados a cabo durante los últimos quince años con ratones genéticamente modificados han permitido identificar nuevos genes de la obesidad. La mayoría de los animales genéticamente manipulados que desarrollan obesidad lo hacen porque comen más. Su obesidad extrema está causada por mutaciones que afectan sus hábitos de alimentación. Es el aumento en la ingesta de alimentos, y no la capacidad de metabolizar la grasa, lo que causa la obesidad. Es decir, la mayoría de los genes de la obesidad regulan el apetito, no el metabolismo lipídico. Así, un ratón mutante obeso, que ha recibido el nombre de ratón *diabético*, es deficiente en el receptor de leptina. El problema de este ratón es que la falta de un receptor de leptina biológicamente activo hace imposible que la leptina pueda informar a las células del cerebro de que hay que parar de comer. Otro ratón obeso, conocido como el ratón *amarillo*, tiene una mutación que afecta a la ruta de la pro-opiomelanocortina (POMC), que es importante para reducir el apetito. La mutación de estos genes de la obesidad produce la acumulación de grasa en los ratones, con independencia de la existencia de otros factores genéticos o ambientales. En general, la situación en humanos es más compleja y resulta poco frecuente que la obesidad se produzca como consecuencia de la mutación de un único gen. En humanos, lo más frecuente es que haya más de un gen implicado en el desarrollo de la obesidad y que el medio ambiente desempeñe también un importante papel en la acumulación de grasa en el cuerpo. Es decir, a nivel individual los hábitos de alimentación son la principal causa que llevan a un individuo a ser obeso, aunque las características genéticas del individuo pueden jugar un papel importante en ese comportamiento.

Los ácidos grasos son los componentes grasos de las moléculas de triglicéridos. Los triglicéridos son el principal componente de la grasa en el tejido adiposo y en los alimentos. Además de provenir de los alimentos, los ácidos grasos también pueden ser sintetizados a partir de los hidratos de carbono, principalmente en el hígado. Los ácidos grasos son una importante reserva de energía, no sólo porque tienen una mayor densidad calórica por gramo que los azúcares o las proteínas, sino porque son hidrófobos, es decir, no atraen agua sino que la repelen y, por consiguiente, pueden almacenarse en el cuerpo de manera más compacta que los carbohidratos o las proteínas, que si atraen agua. Así, mientras que la densidad calórica de la grasa es de 9 kcal por gramo, la de los azúcares y las proteínas es de alrededor de 4 kcal por gramo. Pero, además, mientras que las grasas no acumulan agua, los carbohidratos acumulan 2 gramos de agua por gramo de azúcar. Es decir, hay aproximadamente seis veces más energía almacenada por gramo de grasa que por gramo de azúcar. O dicho de otra manera, si el cuerpo humano almacenase energía en forma de carbohidratos en lugar de en forma de grasa, un individuo necesitaría almacenar cerca de 100 kilos de glucógeno para tener la energía equivalente a 15

kilos de grasa, que es, aproximadamente, la cantidad de grasa que tiene un individuo adulto no obeso.

Cuando los triglicéridos de los alimentos entran en el estómago, los ácidos estomacales, las sales biliares y las enzimas digestivas, conocidas como lipasas, rompen estos triglicéridos en sus dos componentes: ácidos grasos y glicerol. Las lipasas son sintetizadas por el páncreas y las sales biliares provienen del hígado a través de la vesícula biliar. Una vez liberados de los triglicéridos, los ácidos grasos entran dentro de las células que forman la pared intestinal para ser convertidos, de nuevo, en triglicéridos y formar, junto a ésteres de colesterol, fosfolípidos y proteínas unas nano-partículas denominadas quilomicrones. Los quilomicrones son transportados a la sangre a través del sistema linfático. Una vez en la sangre, los quilomicrones intercambian triglicéridos por ésteres de colesterol con otras lipoproteínas denominadas lipoproteínas de alta densidad. A su paso, a través de los capilares sanguíneos, por el tejido adiposo, los músculos, el corazón y otros tejidos no hepáticos, los quilomicrones van perdiendo su carga de ácidos grasos mediante la acción de la enzima lipoproteína lipasa. Los ácidos grasos así generados se oxidan para producir la energía que cada uno de estos tejidos requiere para cumplir su función biológica, o se acumulan en forma de triglicéridos. Finalmente, el remanente de quilomicrones en la sangre, prácticamente liberados totalmente de su carga de triglicéridos, es transportado al interior de las células del hígado para ser metabolizados.

La ingesta de alimentos también induce la secreción de insulina. La secreción de insulina por las células b del páncreas estimula la síntesis de glucógeno en el músculo y en el hígado. En el tejido adiposo, la insulina también estimula el metabolismo de la glucosa y la síntesis de glicerol, la molécula a la que se unen los ácidos grasos para formar triglicéridos. También en el hígado, la insulina suprime la gluconeogénesis –síntesis de glucosa y glucógeno– y acelera la glicolisis –metabolismo de la glucosa–, lo cual aumenta la síntesis de ácidos grasos que se acumulan en forma de triglicéridos. Si de forma crónica la ingesta de grasa e hidratos de carbono es superior a su consumo, el exceso de energía se acumula en forma de triglicéridos, principalmente en el tejido adiposo, causando obesidad. Parte de este exceso de grasa acumulado en el tejido adiposo es transportado hacia el hígado, a través de la sangre, en forma de ácidos grasos libres unidos a la albúmina. Finalmente, estos ácidos grasos se acumulan en forma de triglicéridos en el hígado, produciendo NAFLD.

Sabemos, desde hace por lo menos 500 años, que cuando se sobrealimentan patos y ocas desarrollan hígado graso. En 1570, Bartolomé Scappi, cocinero del papa Pío V, publicó un libro de cocina titulado *Opera* en el que escribía que «el hígado de las ocas domésticas criadas por lo judíos alcanzaba un tamaño extremo de 3 libras». La sobrealimentación no sólo produce NAFLD en las aves. En el laboratorio, la sobrealimentación de ratas y ratones con

una dieta rica en ácidos grasos y carbohidratos, para inducir la generación de hígado graso sigue siendo un método experimental de uso muy extendido.

Las investigaciones llevadas a cabo durante los últimos diez años empleando ratones genéticamente modificados han sido fundamentales para demostrar que, cuando se inactivan ciertas enzimas que son clave para la síntesis hepática de los ácidos grasos y triglicéridos, tales como la acetil-coenzima A carboxilasa, diacilglicerol aciltransferasa, elongasa de ácidos grasos de cadena larga, glicerol 3-fosfato aciltransferasa mitocondrial y esteroil-coenzima A desaturasa, se previene la formación de hígado graso inducida por una dieta rica en grasa y carbohidratos (Postic y Girard 2008). Estos datos sugieren que la disminución de la síntesis hepática de triglicéridos es, potencialmente, una importante diana terapéutica para el tratamiento de NAFLD. Sin embargo, es importante hacer hincapié en que la acumulación de triglicéridos en el hígado no es necesariamente tóxica, sino que es quizás una forma de proteger al hígado de la toxicidad causada por los ácidos grasos libres, es decir, los ácidos grasos que no están unidos a moléculas de glicerol formando triglicéridos. Por ejemplo, en los ratones *obesos* la inhibición de la síntesis de triglicéridos mejora la esteatosis pero empeora el daño hepático –necrosis, inflamación y fibrosis– (Yamaguchi y cols. 2007). Los ácidos grasos libres, si no se oxidan para producir energía, son metabolizados por el sistema microsomal denominado citocromo P450 2E1 –CYP2E1 es particularmente activo en el hígado y, además de metabolizar sustancias exógenas como el alcohol, drogas y pro-carcinógenos, también participa en el metabolismo del colesterol, ácidos biliares y ácidos grasos–. El metabolismo de los ácidos grasos a través de CYP2E1 genera sustancias citotóxicas, como las sustancias reactivas del oxígeno (ROS) y lípidos peroxidados, que producen inflamación y necrosis hepática.

En el hígado, las moléculas de triglicéridos se acumulan en el citoplasma de los hepatocitos formando pequeñas gotas de lípidos. Estas gotas de lípidos no son una simple acumulación de triglicéridos, como las que se forman cuando se mezcla aceite con agua, sino que son organelos que requieren, para formarse, de la presencia de ciertas proteínas específicas. Una de estas proteínas recibe el nombre de ADFP. Los ratones deficientes en ADFP no desarrollan NAFLD cuando se sobrealimentan con una dieta rica en grasa (Chang y cols 2006). Aunque ADFP es potencialmente una diana terapéutica para el tratamiento de NAFLD, se desconoce si la inhibición de la acumulación de triglicéridos en animales *obesos* mediante la inhibición de ADFP puede aumentar el daño hepático. Otra aproximación experimental, que ha sido empleada para prevenir NAFLD, es bloquear la actividad de ciertos factores de transcripción –proteínas que se unen al DNA y que regulan la expresión de genes específicos– que controlan la síntesis de lípidos. Uno de estos factores de transcripción,

que recibe el nombre de SREBP-1c, media el efecto de la insulina sobre la expresión de las enzimas que regulan la síntesis de ácidos grasos. Los ratones *obesos* deficientes en SREBP-1c mejoran la esteatosis (Yahagi y cols. 2002), aunque se desconoce si a largo plazo la inhibición de SREBP-1c puede aumentar el daño hepático. En resumen, aunque la inhibición de la síntesis de triglicéridos o de su acumulación en forma de vesículas en el hígado son, teóricamente, buenas aproximaciones terapéuticas para prevenir NAFLD, es importante recordar que estos procedimientos no están exentos de posibles efectos secundarios que pueden producir daño hepático y, por consiguiente, su aplicación clínica no es obvia.

Sorprendentemente, la desnutrición puede también provocar hígado graso. El mecanismo por el que la desnutrición provoca NAFLD no es del todo conocido, aunque los estudios realizados en los últimos años, empleando animales genéticamente modificados, han proporcionado nuevos datos sobre la importancia de ciertos nutrientes en el desarrollo de NAFLD.

En 1930, Banting y Best, los descubridores de la insulina, observaron que los perros diabéticos tratados con insulina desarrollaban hígado graso y que esta situación podía ser corregida mediante la administración de colina –la colina es un micro-nutriente precursor de la síntesis de metionina–. Unos años más tarde, Best, Du Vigneaud y otros grupos de investigación observaron que, cuando se alimenta a ratas o ratones con una dieta deficiente en metionina y colina, en unas pocas semanas también desarrollan esteatosis que progresa a NASH, e incluso en algunos animales a HCC, si esta dieta se mantiene. Estos animales alimentados con una dieta deficiente en metionina y colina no sólo no son *obesos* sino que, en general, pesan menos que los ratones alimentados con una dieta normal. Estos experimentos, además de relacionar la esteatosis con la diabetes, proporcionaron evidencia, por primera vez, sobre la importancia de un grupo de micronutrientes denominados donantes de grupos metilo –colina, metionina, betaina, y ácido fólico– en la prevención de la esteatosis (Mato y cols. 2008).

En mamíferos, incluidos los humanos, la metionina es un aminoácido esencial, es decir, no puede ser sintetizado por el cuerpo sino que tiene que ser suministrado con los alimentos. Cuando se administra metionina a una persona por vía oral, los niveles en sangre de este aminoácido aumentan de manera transitoria recuperando los niveles basales en dos o tres horas. La velocidad a la que una persona recupera los niveles basales de metionina después de su ingesta es un indicador del metabolismo de este aminoácido por el cuerpo. En pacientes cirróticos, el metabolismo de la metionina es marcadamente más lento que en individuos con función hepática normal. El primer paso en el metabolismo de la metionina es su conversión a S-adenosilmetionina –S-AMe–, una molécula descubierta por Giulio Cantoni en 1953 (Cantoni 1975). La S-AMe ocupa

un lugar especial en la biología debido a la capacidad que tiene de modificar, mediante la adición de un grupo metilo –se denomina grupo metilo a un átomo de carbono que tiene unidos tres átomos de hidrógeno–, a otras moléculas, tales como el DNA, las proteínas y los fosfolípidos, modificando su actividad biológica. Esta reacción, conocida con el nombre general de metilación, puede hacer que ciertos genes no se expresen, es decir, puede causar el mismo resultado que una mutación genética pero el mecanismo no es genético sino epigenético.

La síntesis de SAMe se encuentra marcadamente reducida en el hígado de los pacientes cirróticos (Duce y cols. 1988), y el tratamiento con SAMe aumenta la supervivencia en pacientes con cirrosis alcohólica (Mato y cols. 1999), lo cual confirma el importante papel de la alteración del metabolismo de la metionina en la progresión de la enfermedad hepática. Consecuentemente, los ratones deficientes en la síntesis hepática de SAMe, aunque no son obesos y tienen un tamaño normal, desarrollan esteatosis, NASH y HCC (Lu y cols. 2001). En los ratones deficientes en la enzima glicina N-metiltransferasa, la principal enzima que metaboliza SAMe en el hígado, la concentración hepática de SAMe es alrededor de 40 veces más elevada que en los ratones normales (Martínez-Chantar y cols. 2008). Sorprendentemente, estos ratones «super-SAMe», aunque tienen un tamaño normal y no son obesos, también desarrollan esteatosis, NASH y HCC. Estos resultados indican que, tanto la deficiencia como el exceso hepático de SAMe inducen NAFLD, e incluso la aparición de HCC, en ausencia de obesidad, lo cual pone de manifiesto la importancia del metabolismo de los grupos metilo en la regulación de la función hepática, y complican la utilización terapéutica de esta molécula.

La actividad CYP2E1 hepática está aumentada en los pacientes con NASH, en los diabéticos y en los individuos que han ayunado prolongadamente. La actividad CYP2E1 hepática también está aumentada en los pacientes con esteatohepatitis alcohólica, una enfermedad muy similar a NASH. Además, la actividad CYP2E1 hepática está incrementada en los animales que han sido alimentados con una dieta deficiente en metionina y colina, y en los ratones deficientes en la síntesis hepática de SAMe. Estos y otros resultados han demostrado la importancia del estrés oxidativo generado por la peroxidación de lípidos vía CYP2E1

en la patogénesis de NASH, es decir, en la progresión de esteatosis a NASH. Sorprendentemente, la deficiencia de CYP2E1 en ratones no previno el desarrollo de NASH inducido por una dieta deficiente en metionina y colina, ni evitó la peroxidación de lípidos, indicando la existencia de un sistema de peroxidación lipídica alternativo que actúa en ausencia de CYP2E1 (Leclercq y cols. 2000). Estos autores también observaron que en los ratones deficientes en CYP2E1, tratados con una dieta deficiente en metionina y colina, la expresión hepática de CYP4A10 y CYP4A14 está inducida y que estas dos enzimas son las responsables de la peroxidación lipídica y generación de ROS en estos animales. CYP4A10 y CYP4A14 forman parte de la familia de enzimas microsomales conocida con el nombre genérico de CYP 450 y de la que CYP2E1 es también uno de sus miembros. Es decir, otros miembros de la familia CYP 450, que en condiciones normales son poco activos, pueden sustituir a CYP2E1 en la peroxidación de lípidos cuando la actividad de esta enzima se encuentra inhibida o mutada. Esto es lo que les ocurre a los ratones «super-SAMe». SAMe es un inhibidor de la expresión hepática de CYP2E1 y, por consiguiente, su expresión está inhibida en los ratones «super-SAMe» aunque hayan desarrollado NAFLD. En estos ratones, como en los animales deficientes en CYP2E1 alimentados con una dieta deficiente en metionina y colina, es la expresión de CYP4A10 y CYP4A14 la que se encuentra estimulada y cataliza la peroxidación de lípidos y formación de ROS.

Una conclusión importante de estos estudios es que las aproximaciones terapéuticas que tienen como diana una única enzima del sistema microsomal CYP P450 no son eficaces para prevenir la generación de ROS y la peroxidación de lípidos y, por consiguiente, bloquear la iniciación y progresión de NASH. Una de las características principales de la biología es la redundancia de las rutas bioquímicas que controlan funciones biológicas esenciales, como la proliferación celular o la defensa contra agentes externos citotóxicos. La ventaja evolutiva de haber desarrollado un sistema complejo, como el CYP 450, integrado por decenas de enzimas cuya misión es proteger al hígado de la acción citotóxica de innumerables xenobióticos, es obvia; a cambio, la redundancia de las enzimas del complejo CYP 450 es una desventaja cuando se persigue neutralizar este sistema y así evitar sus efectos secundarios, como es la progresión a NASH en individuos con esteatosis.

Bibliografía

- Abdelmalek M. F. y A. M. Diehl. «Nonalcoholic fatty liver disease as a complication of insulin resistance». *Journal Med. Clin. North Am* 91 (noviembre 2007): 1.125.
- Adams L. A. y K. D. Lindor. «Nonalcoholic fatty liver disease». *Journal Ann. Epidemiol* 17 (noviembre 2007): 863.
- Ahmed M. H. y Byrne C. D. «Modulation of sterol regulatory element binding proteins (SREBPs) as potential treatments for non-alcoholic fatty liver disease (NAFLD)». *Journal Drug Discovery Today* 12 (septiembre 2007): 740.
- «Alzheimer». *Psych. genchtl Med* 64 (1907): 146.
- Amadoro, G., M. T. Ciotti, M. Costanzi, V. Cestari, P. Calissano y N. Canu. «NMDA receptor mediates tau-induced neurotoxicity by calpain and ERK/MAPK activation». *Journal Proc Natl Acad Sci USA* 103 (febrero 2006): 2892.
- Angulo P. «Obesity and nonalcoholic fatty liver disease». *Journal Nutr. Rev.* 65 (junio 2007): S57.
- Arriagada, P. V., J. H. Growdon, E. T. Hedley-Whyte y B. T. Hyman. «Neurofibrillary tangles but not senile plaques parallel duration and severity of Alzheimer's disease». *Journal Neurology* 42 (marzo 1992): 631.
- Avila, J., J. J. Lucas, M. Pérez y F. Hernández. «Role of tau protein in both physiological and pathological conditions». *Journal Physiol Rev.* 84 (abril 2004): 361.
- Baki, L., J. Shioi, P. Wen, Z. Shao, A. Schwarzman, M. Gama-Sosa, R. Neve y N. K. Robakis. «PS1 activates PI3K thus inhibiting GSK-3 activity and tau overphosphorylation: effects of FAD mutations». *Journal Embo* 23 (julio 2004): 2.586.
- Blennow, K. y H. Zetterberg. «Pinpointing plaques with PIB». *Journal Nat Med* 12 (julio 2006): 753.
- Braak, H. y E. Braak. «Neuropathological staging of Alzheimer-related changes». *Journal Acta Neuropathol* 82 (1991): 239.
- Brittberg, M., A. Lindahl, A. Nilsson, C. Ohlsson, O. Isaksson y L. Peterson. «Treatment of deep cartilage defects in the knee with autologous chondrocyte transplantation». *Journal N Engl J Med* 331 (octubre 1994): 889.
- Cantoni G.L. «Biochemical methylations: selected aspects». *Journal Ann. Rev. Biochem.* 45 (1975): 285-306.
- Chang B. H.-J., J. Li, A. Paul, S. Taniguchi, V. Nanegari, W. C. Herid y L. Chan. «Protection against fatty liver but normal adipogenesis in mice lacking adipose differentiation-related protein». *Journal Mol. Cell. Biol.* 26 (febrero 2006): 1.063.
- Duce A. M., P. Ortiz, C. Cabrero y J. M. Mato. «S-Adenosyl-L-methionine synthetase and phospholipid methyltransferase are inhibited in human cirrhosis». *Journal Hepatology* 8 (enero-febrero 1988): 1.530.
- Engel, T., J. J. Lucas, P. Gómez-Ramos, M. A. Morán, J. Avila y F. Hernández. «Coexpression of FTDP-17 tau and GSK-3beta in transgenic mice induce tau polymerization and neurodegeneration». *Journal Neurobiol Aging* 27 (septiembre 2006): 1.258.
- Fire, A., S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver y C. C. Mello. «Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*». *Journal Nature* 391 (febrero 1998): 806.
- Glenner, G. G. y C. W. Wong. «Alzheimer's disease and Down's syndrome: sharing of a unique cerebrovascular amyloid fibril protein». *Journal Biochem Biophys Res Commun* 122 (agosto 1984): 1.131.
- Gómez-Ramos, A., M. Díaz-Hernández, A. Rubio, M. T. Miras-Portugal y J. Avila. «Extracellular tau promotes intracellular calcium increase through M1 and M3 muscarinic receptors in neuronal cells». *Journal Mol Cell Neurosci* 37 (abril 2008): 673.
- Grundke-Iqbal, I., K. Iqbal, Y. C. Tung, M. Quinlan, H. M. Wisniewski y L. I. Binder. «Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology». *Journal Proc Natl Acad Sci USA* 83 (julio 1986): 4.913.
- Hardy, J. y D. J. Selkoe. «The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics». *Journal Science* 297 (julio 2002): 353.
- Harkany, T., I. Abraham, W. Timmerman, G. Laskay, B. Toth, M. Sasvari, C. Konya, et al. «Beta-amyloid neurotoxicity is mediated by a glutamate-triggered excitotoxic cascade in rat nucleus basalis». *Journal Eur J Neurosci* 12 (agosto 2000): 2.735.
- Hernández, F., J. Borrell, C. Guaza, J. Avila y J. J. Lucas. «Spatial learning deficit in transgenic mice that conditionally over-express GSK-3beta in the brain but do not form tau filaments». *Journal J Neurochem* 83 (diciembre 2002): 1.529.
- Kennedy, D. «Breakthrough of the year». *Journal Science* 318 (diciembre 2007): 1.833.
- Kerr, J. N. y W. Denk. «Imaging in vivo: watching the brain in action». *Journal Nat Rev Neurosci* 9 (marzo 2008): 195.
- Kukar, T. L., T. B. Ladd, M. A. Bann, P. C. Fraering, R. Narlawar, G. M. Maharvi, B. Healy, et al. «Substrate-targeting gamma-secretase modulators». *Journal Nature* 453 (junio 2008): 925.
- Kurtzberg, J., M. Laughlin, M. L. Graham, C. Smith, J. F. Olson, E. C. Halperin, G. Ciocci, et al. «Placental blood as a source of hematopoietic stem cells for transplantation into unrelated recipients». *Journal N Engl J Med* 335 (julio 1996): 157.
- Leclercq I. A., G. C. Farrell, J. Field, D. R. Bell, F. J. González y G. H. Robertson. «CYP2E1 and CYP4A as microsomal catalysts of lipid peroxides in murine nonalcoholic steatohepatitis». *Journal J. Clin. Invest.* 105 (abril 2000): 1.067.
- Logothetis, N. K. «What we can do and what we cannot do with fMRI». *Journal Nature* 453 (junio 2008): 869.
- Lu S. C., L. Álvarez, Z. Z. Huang, L. Chen, W. An, F. J. Corrales, M. A. Avila, et al. «Methionine adenosyltransferase 1A knockout mice are predisposed to liver injury and exhibit increased expression of genes involved in proliferation». *Journal Proc. Natl. Acad. Sci. USA* 98 (mayo 2001): 5.560.
- Lucas, J. J., F. Hernández, P. Gómez-Ramos, M. A. Morán, R. Hen y J. Avila. «Decreased nuclear beta-catenin, tau hyperphosphorylation and neurodegeneration in GSK-3beta conditional transgenic mice». *Journal Embo* 20 (enero 2001): 27.
- Mato J. M., M. L. Martínez-Chantar y S. C. Lu. «Methionine metabolism and liver disease». *Journal Annu. Rev. Nutr.* 28 (agosto 2008): 273.
- Mackman, N. «Triggers, targets and treatments for thrombosis». *Journal Nature* 451 (febrero 2008): 914.
- Mani, S. A., W. Guo, M. J. Liao, E. N. Eaton, A. Ayyanan, A. Y. Zhou, M. Brooks et al. «The epithelial-mesenchymal transition generates cells with properties of stem cells». *Journal Cell* 133 (mayo 2008): 704.
- Martínez-Chantar M. L., M. Vázquez-Chantada, U. Ariz, N. Martínez, M. Varela, Z. Luka, A. Capdevila et al. «Loss of glycine N-methyltransferase gene leads to steatosis and hepatocellular carcinoma in mice». *Journal Hepatology* 47 (abril 2008): 1.191.
- Masters, C. L., G. Simms, N. A. Weinman, G. Multhaup, B. L. McDonald y K. Beyreuther. «Amyloid plaque core protein in Alzheimer disease and Down syndrome». *Journal Proc Natl Acad Sci USA* 82 (junio 1985): 4.245.

- Mato J. M., J. Camara, J. Fernández de Paz, L. Caballería, S. Coll, A. Caballero, L. García-Buey, et al. «S-adenosylmethionine in alcoholic liver cirrosis: a randomized placebo-controlled, double-blind, multicenter clinical trial». *Journal J. Hepatol.* 30 (junio 1999): 1.081.
- Montejo de Garcini, E., L. Serrano y J. Avila. «Self assembly of microtubule associated protein tau into filaments resembling those found in Alzheimer disease». *Journal Biochem Biophys Res Commun* 141 (diciembre 1986): 790.
- Mount, C. y C. Downton. «Alzheimer disease: progress or profit?». *Journal Nat Med* 12 (julio 2006): 780.
- O'Connor, N. E., J. B. Mulliken, S. Banks-Schlegel, O. Kehinde y H. Green. «Grafting of burns with cultured epithelium prepared from autologous epidermal cells». *Journal Lancet* 8.211 (enero 1981): 75.
- Parsons, C. G., W. Danysz y G. Quack. «Memantine is a clinically well tolerated N-methyl-D-aspartate (NMDA) receptor antagonist--a review of preclinical data». *Journal Neuropharmacology* 38 (junio 1999): 735.
- Pedersen, W. A., P. J. McMillan, J. J. Kulstad, J. B. Leverenz, S. Craft y G. R. Haynatzki. «Rosiglitazone attenuates learning and memory deficits in Tg2576 Alzheimer mice». *Journal Exp Neurol* 199 (junio 2006): 265.
- Postic, C. y J. Girard. «Contribution of de novo fatty acid sintesis to hepatic steatosis and insulin resistance: lessons from genetically engineered mice». *Journal J. Clin. Invest.* 118 (marzo 2008): 829.
- Price, D. L., R. E. Tanzi, D. R. Borchelt y S. S. Sisodia. «Alzheimer's disease: genetic studies and transgenic models». *Journal Annu Rev Genet* 32 (1998): 461.
- Rader, D. J. y A. Daugherty. «Translating molecular discoveries into new therapies for atherosclerosis». *Journal Nature* 451 (febrero 2008): 904.
- Ray, S., M. Britschgi, C. Herbert, Y. Takeda-Uchimura, A. Boxer, K. Blennow, L. F. Friedman et al. «Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins». *Journal Nat Med* 13 (noviembre 2007): 1.359.
- Reisberg, B., R. Doody, A. Stoffer, F. Schmitt, S. Ferris y H. J. Mobius. «Memantine in moderate-to-severe Alzheimer's disease». *Journal N Engl J Med* 348 (abril 2003): 1.333.
- Segers, V. F. y R. T. Lee. «Stem-cell therapy for cardiac disease». *Journal Nature* 451 (febrero 2008): 937.
- Stephen Salloway, J. M., F. W. Myron, y J. L. Cummings. «Disease-modifying therapies in Alzheimer's disease». *Journal/Alzheimer's Et Dementia* 4 (2008): 65.
- Sturm, R. «The effects of obesity, smoking and drinking on medical problems and costs». *Journal Health Affaire* 21 (marzo-abril 2002): 245.
- Takahashi, K. y S. Yamanaka. «Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors». *Journal Cell* 126 (agosto 2006): 663.
- Thomson, J. A., J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall y J. M. Jones. «Embryonic stem cell lines derived from human blastocysts». *Journal Science* 282 (noviembre 1998): 1.145.
- Vas, C. J. «Alzheimer's disease: The brain killer». World Health Organization (2001).
- Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith et al. «The sequence of the human genome». *Journal Science* 291 (febrero 2001): 1.304.
- Wang, J., L. Ho, W. Zhao, K. Ono, C. Rosensweig, L. Chen, N. Humala, D. B. Teplow y G. M. Pasinetti. «Grape-derived polyphenolics prevent Abeta oligomerization and attenuate cognitive deterioration in a mouse model of Alzheimer's disease». *Journal J Neurosci* 28 (junio 2008): 6.388.
- Wernig, M., J. P. Zhao, J. Pruszak, E. Hedlund, D. Fu, F. Soldner, V. Broccoli et al. «Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease». *Journal Proc Natl Acad Sci U S A* 105 (abril 2008): 5.856.
- Wold, B. y R. M. Myers. «Sequence census methods for functional genomics». *Journal Nat Methods* 5 (enero 2008): 19.
- Yahagi N., H. Shimano, A. H. Hasty, T. Matsuzaka, T. Ide, T. Yoshikawa, M. Amemiya-Kudo et al. «Absence of sterol regulatory element-binding protein-1 (SREBP-1) ameliorates fatty livers but not obesity or insulin resistance in Lep (ob/Lep(ob) mice». *Journal J. Biol. Chem.* 277 (mayo 2002): 19.353.
- Yamaguchi K., L. Yang, S. McCall, J. Huang, X. X. Yu, S. K. Pandey, S. Bhanot et al. «Inhibiting triglyceride synthesis improves hepatic steatosis but exacerbates liver damage and fibrosis in obese mice with nonalcoholic steatohepatitis». *Journal Hepatology* 45 (junio 2007): 1.366.
- Zhang Y., M. Proenca, M. Maffei, M. Barone, L. Leopold y J. M. Friedman. «Positional cloning of the mouse gene obese and its human homologue». *Journal Nature* 372 (diciembre 1994): 425.

clonación de mamíferos: algo más que una simple oveja

ALEXANDER KIND Y ANGELIKA SCHNIEKE

Desde que en 1997 se hiciera público el caso de Dolly, la clonación por transferencia nuclear ha despertado un enorme interés en los medios de comunicación. Pero los comentaristas se han preocupado más de especular sobre posibles ejércitos de dictadores o deportistas clonados, y aplicaciones secundarias, como la sustitución de mascotas. En este ensayo queremos situar la transferencia nuclear en un contexto más amplio y destacar las consecuencias más sutiles y a la vez profundas del trabajo.

Así que empezamos por preguntarnos por qué surgió una forma tan complicada de producir animales. Quienes visiten Escocia pronto se darán cuenta de que no hay escasez de ovejas. De hecho, tras el experimento de Dolly se esconden dos motivos. Uno era estrictamente comercial y consistía en desarrollar una herramienta para la producción rápida de animales idénticos para la biotecnología. El segundo motivo, mucho más profundo, era la simple curiosidad científica y una oportunidad de tratar un dilema biológico muy antiguo. En su calidad de animales complejos, las ranas, los ratones, las ovejas y los seres humanos proceden de una única célula a partir de la cual se forman muchos tipos diferentes de células. ¿Cómo alcanzan sus destinos y cómo conservan o cambian de identidad?

Primeras investigaciones y principios fundamentales

Los expertos han debatido la cuestión del desarrollo animal desde la Antigüedad. En el siglo III a.C., Aristóteles reconoció la importancia de la reproducción sexual y propuso dos modelos alternativos. O la estructura del animal completo ya estaba preformada en miniatura dentro del óvulo o el embrión, o nuevas estructuras iban surgiendo poco a poco. Aristóteles se inclinó por la segunda idea, pero al carecer de la tecnología adecuada la cuestión siguió siendo durante siglos objeto de interminables debates filosóficos. El preformismo se convirtió en la doctrina más extendida en la Europa de los siglos XVII y XVIII, tal y como ilustra el grabado del siglo XVII de la figura 1. Animado por el descubrimiento de los espermatozoides, o «animáculos» como se les llamaba entonces, el físico y microscopista Nicholas Hartsoeker propuso que en su interior podía estar la estructura de un diminuto feto. Hartsoeker conjeturó que la cabeza del espermatozoide crecía hasta formar el feto y que la cola se transformaba en el cordón umbilical, mientras la función del óvulo sólo consistía en proporcionar un nido para facilitar el desarrollo del nuevo ser.

Pero hasta después de 1830 no se pudieron realizar estudios fiables. Fue entonces cuando el naturalista bri-

tánico aficionado Joseph Jackson Lister inventó el microscopio compuesto. Los instrumentos con más de una lente proporcionaban suficiente resolución para distinguir por primera vez la estructura detallada de tejido vivo. Podría decirse que la biología moderna nació en 1839, cuando Theodor Schwann y Matthias Schleiden demostraron que los seres vivos se componían de células. Poco después, Albrecht von Kölliker demostró que los espermatozoides y los ovocitos (huevos) también son células, pero la forma en que interactuaban para constituir un nuevo organismo seguía siendo un misterio. El eminente químico Justus von Liebig sugirió que el espermatozoide transmite sus cualidades masculinas al ovocito a través de la vibración energética de sus colas. En 1854, George Newport describió sus estudios de fertilización en ranas y sugirió que el espermatozoide realiza su aportación penetrando el huevo. Hacia la misma fecha, las investigaciones microscópicas revelaron que nuevas células surgían de la división del huevo fertilizado, por lo que resultaba inverosímil que el desarrollo se produjera por preformación.

A Oskar Hertwig se le atribuyen los inicios del estudio sobre fertilización y desarrollo embrionario temprano del erizo de mar, un campo muy productivo que proporcionó gran parte de la información que se aplicó posteriormente a otras especies. Los erizos de mar son idóneos para los estudios microscópicos porque sus huevos son muy claros. En 1876, Hertwig describió sus observaciones tras añadir

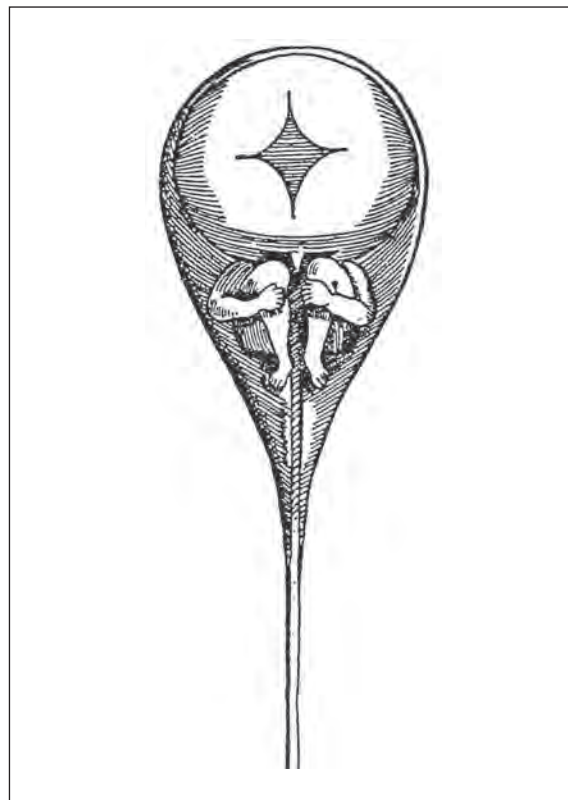


Figura 1. Feto preformado dentro de una cabeza de zvvv. N. Hartsoeker (1694). *Essai de dioptrique*, Paris.

esperma a los huevos. En particular, observó la presencia de dos núcleos dentro del huevo, uno de los cuales provenía del espermatozoide, e indicó cómo se fusionaban. Ésta fue la primera explicación del papel de los dos padres en la reproducción. También destacó la importancia del núcleo, y de los cuerpos coloreados en su interior que pudieron apreciarse al utilizar los tintes de anilina y que en la década de 1880 pasaron a llamarse «cromosomas».

Se puede decir que el biólogo alemán August Weismann va inmediatamente detrás de Charles Darwin en cuanto a aportaciones a la biología teórica. En 1892 Weismann propuso una idea atrevida: los núcleos de los ovocitos y del espermatozoide contenían una sustancia hereditaria, que además constituía la única continuidad orgánica entre generaciones (Weismann 1892). Este principio sentó los cimientos de toda la genética y la biología evolutiva. La teoría del «plasma germinal» de Weismann establece que las células germinales tienen un linaje muy distinto del resto de las células del cuerpo, las células somáticas, y que las características que el cuerpo adquiere a lo largo de la vida no se transmiten a las células germinales. Esto contradecía de forma explícita las teorías de Jean-Baptiste Lamarck, muy aceptadas en la época incluso por Darwin. En los siguientes 20 años, la corriente de pensamiento iniciada por Weismann pasó a convertirse en la ciencia moderna sobre genética y desarrollo. En 1900 se redescubrieron los trabajos de Gregor Mendel sobre híbridos de guisantes y con ellos su concepto de segregación de caracteres independientes. Dos años más tarde Theodor Boveri y Walter Sutton habían descubierto que los elementos que determinan los caracteres identificados por Mendel se encuentran en los cromosomas. En 1907, Boveri demostró que la presencia de un juego de cromosomas normal es necesaria para el desarrollo embrionario del erizo de mar. En 1915, Thomas Morgan descubrió la localización física de los genes en los cromosomas de la mosca de la fruta en su obra maestra *The Mechanism of Mendelian Heredity* (*El mecanismo de la herencia mendeliana*) (Morgan et al. 1915).

En la actualidad, estas ideas conforman la base de la biología, pero durante el siglo xx se consideraron «degeneradas y fascistas». Los ideólogos de la Rusia soviética rechazaron violentamente la teoría del plasma germinal. A partir de los primeros años de la década de 1930 hasta 1964, la política soviética oficial rechazó las ideas de Weismann y la genética en su totalidad. Desde luego, Stalin no será recordado por su interés en la biología del desarrollo y todo indica que actuó por motivos políticos. La herencia de caracteres adquiridos hacía posible que la raza humana se perfeccionara gracias a la política «materialista progresista». De esta forma, los líderes soviéticos podían justificar las penurias que tenía que soportar la gente al considerarlas necesarias para la producción de futuras generaciones de perfectos comunistas. El ambiente político también favoreció el ascenso del conocido agrónomo soviético Trofim Lyenko. En un acalorado dis-

curso en la Academia de Ciencias Agrícolas pronunciado en agosto 1948, Lysenko denunció extensamente a Weismann y se burló de la «seudociencia burguesa» de sus seguidores. Sin ser consciente de ello, Lysenko realizó una descripción bastante acertada del concepto del genoma y de una sustancia que acabaría llamándose ADN:

Weismann negó la capacidad de heredar los caracteres adquiridos y elaboró una idea sobre una sustancia hereditaria especial que debía buscarse en el núcleo. Afirmó que la herencia estaba contenida en el cromosoma. [...] Una sustancia hereditaria e inmortal, independiente de los rasgos cualitativos presentes en el desarrollo del cuerpo vivo, y que permanece en el cuerpo mortal [...] éste es el concepto denodadamente idealista y profundamente místico de Weismann (Lysenko 1948).

Pero la naturaleza se muestra brutalmente indiferente a la teoría política. Los métodos de Lysenko fueron responsables de repetidos fracasos de cosechas de la URSS y, cuando China adoptó políticas agrícolas similares en 1958 en la época del «Gran Salto Adelante», contribuyeron a la mayor hambruna conocida de la historia, la de 1959-1961.

Clonación e identificación de las células

Junto con su concepto de plasma germinal, Weismann propuso la primera teoría experimental del desarrollo animal, un proceso denominado desarrollo en mosaicos. Afirmó que la célula única embrionaria, el cigoto, contiene factores o determinantes localizados en regiones diferenciadas. Al escindirse, los determinantes se distribuyen de forma desigual en células hijas y controlan su desarrollo futuro. El proceso continúa mientras los diferentes tipos de células se forman por «diferenciación», a medida que el embrión se desarrolla. Este modelo predice de forma clara que las células individuales del embrión en desarrollo no deberían compartir el mismo potencial. Sin embargo, en 1982, Hans Driesch proporcionó la primera prueba que desmentía la teoría de Weismann (Driesch 1892). Las células de los embriones tempranos de erizos de mar podían separarse y formar cada uno de ellos una larva completa. La división en la fase bicelular llevaba a la formación de dos larvas normales y las células individuales de la fase tetracelular producían cuatro larvas normales. De hecho, éstos fueron

los primeros animales clonados de forma experimental.

En un discurso pronunciado en la Universidad de Londres en octubre de 1913, Driesch afirmó que el embrión es un «sistema equipotencial armónico [...] donde cada elemento es capaz de desempeñar diferentes funciones. La verdadera función que desempeña en cada caso en particular depende de su posición». Desde entonces se han realizado muchas demostraciones que confirman que los embriones de muchos vertebrados, incluyendo los mamíferos, pueden reorganizarse al cambiar la configuración del número de células, para después recuperarse y formar un animal completo normal.

Al principio, la clonación por transferencia nuclear se presentó como un método adicional para probar que los núcleos de las células embrionarias tempranas y adultas tenían un potencial de desarrollo equivalente, y es una idea más antigua de lo que se cree. Yves Delage, un olvidado biólogo marino francés, hizo la primera referencia a este procedimiento en 1895, arguyendo que «siempre que no se produzcan daños, si el núcleo del óvulo puede sustituirse con el núcleo de una célula embrionaria común, entonces ese óvulo podría desarrollarse sin cambios» (Beetschen y Fischer 2004). Sin embargo, no se sabe si Delage llegó a completar este experimento. Este honor suele recaer en Hans Spemann, un antiguo alumno de Boveri. En 1928 Spemann realizó la primera transferencia nuclear en un notable estudio de microcirugía (Spemann 1928). La figura 2 muestra los dibujos del propio Spemann. Con ayuda de unas micropinzas y un mechón de pelo de su hija pequeña, escindió un embrión unicelular de salamandra en dos partes, una de las cuales contenía el núcleo de la célula (fig. 2A). Al desarrollarse, esta parte se dividió y formó un embrión, mientras que la otra permaneció como una bolsa clara de citoplasma (figs. 2B y 2C). El embrión siguió desarrollándose hasta alcanzar la fase de 16 células. En ese momento, se devolvió un solo núcleo al citoplasma vacío (fig. 2D). Esta célula única se convirtió en un embrión de salamandra normal en una fase ligeramente anterior (fig. 2E). Esto puso de manifiesto que los núcleos de células embrionarias eran capaces de formar un animal completo.

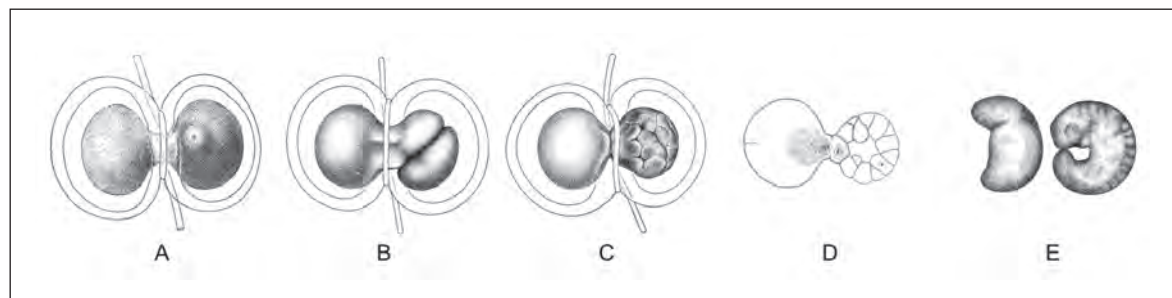


Figura 2. El experimento de transferencia nuclear de Hans Spemann con huevos de salamandra y un pelo de recién nacido. A) Se hizo un nudo para dividir un embrión unicelular en dos mitades conectadas por un puente de citoplasma, con el núcleo en la mitad derecha. B) Primera división celular en la mitad derecha nucleada. C) Fase de desarrollo posterior en la mitad nucleada. D) Durante la división celular, un núcleo hijo pasa a través del citoplasma vacío. E) Ahora el desarrollo lo convierte en dos mitades, pero con un retraso temporal, el embrión de la izquierda es más joven que su gemelo. H. Spemann (1936).

No se sabe si Spemann conocía el estudio anterior de Delage, pero en 1936, ya jubilado, Spemann propuso lo que llamó «un experimento de clonación propio del mundo de la fantasía» (Spemann 1936). Si se podían transferir los núcleos de las células en fases más avanzadas de desarrollo para devolverlos a huevos fertilizados, sería factible identificar de forma sistemática el momento en el que las células conservan o pierden su capacidad para formar un organismo completo, una cualidad que en la actualidad se denomina «totipotencia celular».

En las siguientes décadas un gran número de biólogos del desarrollo centraron sus estudios en erizos de mar y anfibios ya que eran relativamente fáciles de criar y manipular. Los ovocitos de rana son células de gran tamaño, de 1 a 2 milímetros de diámetro, y bastante visibles ya que forman esferas de color gris oscuro o marrón dentro de una capa protectora de gelatina. A principios de la década de 1950, Robert Briggs y Thomas King pusieron en práctica el experimento «fantasioso» de Spemann con ranas (Briggs y King 1952). Extrajeron el núcleo de un ovocito activado valiéndose de una aguja de cristal. Después, una célula única diseccionada a partir de un embrión en una fase posterior se introdujo en una pipeta de cristal muy fina conectada por un tubo de goma a una jeringa. La célula se rompió al introducirse en la pipeta y el núcleo liberado se inyectó en el huevo enucleado. Tras hacer un cultivo con los embriones reconstruidos, descubrieron que los núcleos celulares de los embriones en fase blastular podían desarrollarse de forma normal para convertirse en larvas en fase de alimentación. Tanto los núcleos como los embriones en una fase más tardía, en los que los linajes de células embrionarias más importantes, como mesodérmicos o endodérmicos, ya estaban establecidos, eran incapaces de hacerlo.

Más tarde John Gurdon y Ron Laskey ampliaron el estudio utilizando núcleos de tejidos juveniles y adultos, como la membrana interdigital de las ranas, y descubrieron que esos animales sobrevivían hasta la fase de renacuajos, pero no mucho más. Gurdon consiguió obtener ranas adultas a partir de tejido intestinal de renacuajos en 1962 (Gurdon 1962), pero la posibilidad de que en su tejido estuvieran presentes células germinales puso en duda los resultados. En aquel momento la consideración más relevante fue que la capacidad de desarrollo de los núcleos trasplantados disminuía con la edad y el ámbito de diferenciación de la célula donante. Los núcleos en la fase más temprana del embrión pueden ser equivalentes, pero en algún momento su destino queda determinado, «activado» por un cambio concreto, como la pérdida o la modificación irreversible del ADN contenido en el núcleo.

Pero esta consideración resulta difícil de conciliar con algunos fenómenos muy conocidos, sobre todo las capacidades regenerativas de la mayoría de los peces y anfibios, como los tritones y las salamandras. Si un tritón pierde una extremidad, las células de los tejidos circundantes

como la piel migran a la herida y realizan un proceso de «desarrollo inverso» diferenciador para formar un blastema. Es un conjunto de células embrionarias que se dividen rápidamente. Las células del blastema se diferencian y se reorganizan para formar una extremidad de sustitución. Esto probaba que algunas células diferenciadas adultas no tienen un destino determinado y pueden cambiar de identidad de forma radical. ¿Era la regeneración de la extremidad tan diferente de la generación de un animal completo mediante transferencia nuclear? ¿O el fracaso de la transferencia nuclear era consecuencia de limitaciones técnicas antes que las biológicas? Estos interrogantes bastaron para que algunos investigadores siguieran experimentando con la determinación celular.

Las ovejas abrieron el camino

Puesto que los biólogos son mamíferos, es lógico que les interesara más investigar especies a las que están próximos que erizos de mar o anfibios. Pero durante muchos años ello entrañaba grandes dificultades técnicas. Los embriones mamíferos crecen dentro de las condiciones controladas del tracto reproductivo femenino y no en un estanque o en agua de mar, y aunque son grandes si los comparamos con otras células de una décima de milímetro aproximadamente, lo cierto es que son prácticamente invisibles sin ayuda del microscopio. Hasta las décadas de 1970 y 1980 los cultivos embrionarios y las técnicas de micro manipulación no mejoraron lo suficiente para que la transferencia de núcleos de mamíferos fuese factible. La idea básica es prácticamente la misma que concibió Spemann: se extrae material genético de un huevo y después se sustituye con el núcleo de otra célula, casi siempre fusionando toda la célula con el ovocito.

Las investigaciones se centraron en el mamífero favorito del laboratorio, el ratón. Sin embargo, los intentos de repetir el experimento de Briggs y Kings con ratones seguían fracasando. En 1981, Karl Illmensee y Peter Hoppe afirmaron que habían clonado ratones mediante la transferencia de núcleos a partir de embriones en fase de blastocito (Illmensee y Hoppe 1981). Sin embargo, más tarde se averiguó y se llegó a la conclusión de que sus conclusiones eran falsas, a pesar de que nunca se probó que se tratara de un fraude intencionado. Después, en 1984, pareció que James McGrath y Davor Solter habían encontrado una solución a la transferencia nuclear en mamíferos. Transfirieron de forma sistemática núcleos de embriones en fase de blastocito de una, dos, cuatro y ocho células a cigotos enucleados, embriones en fase unicelular. Los núcleos de embriones unicelulares lograron desarrollarse y convertirse en blastocitos; el éxito fue mucho menor en los núcleos en la fase bicelular y fracasó estrepitosamente en fases más avanzadas. Lo atribuyeron a una pérdida rápida de totipotencia durante el desarrollo. Su estudio concluye con una afirmación rotunda: «La clonación de mamíferos por transferencia nuclear es biológicamente imposible». (McGrath y Solter 1984)

En retrospectiva, es de lamentar que muchos de los primeros intentos se realizaran con ratones. Desde entonces se ha averiguado que son una de las especies más difíciles de clonar por transferencia nuclear. Por esta razón, por muy extraña que parezca, la mayor parte de los descubrimientos se hicieron utilizando ganado. Los primeros mamíferos clonados por transferencia nuclear fueron tres ovejas de Suffolk. Steen Willadsen fusionó células únicas extraídas de embriones octocelulares con huevos enucleados sin fertilizar (Willadsen 1986). Paradójicamente, estas ovejas nacieron en 1984, tan sólo unos meses antes de que McGrath y Solter afirmaran que la clonación de mamíferos era imposible. El motivo de esta discrepancia era de tipo técnico. McGrath y Solter habían utilizado cigotos enucleados porque los ovocitos de ratón son demasiado frágiles para sobrevivir a la transferencia nuclear. Willadsen

pudo utilizar ovocitos sin fertilizar, que son más resistentes en las ovejas. Desde entonces, años de trabajo han demostrado que en gran número de especies, los ovocitos sin fertilizar pueden recibir con éxito transferencias nucleares, mientras que los cigotos sólo pueden utilizarse en una fase muy concreta. Hemos tenido que esperar hasta este año para tener un modelo que explicara esta diferencia, del que hablaremos más adelante (Egli, Birkhoff y Eggan 2008).

Durante la siguiente década, la transferencia nuclear se llevó a cabo en distintos mamíferos, pero, al igual que ocurrió con la rana, sólo uno tuvo éxito utilizando células obtenidas de embriones muy tempranos, o cultivados durante periodos de tiempo muy cortos.

A principios de la década de 1990, Keith Campbell e Ian Wilmut del Roslin Institute, cerca de Edimburgo, empezaron a estudiar cómo la elección del ovocito receptor y la fase

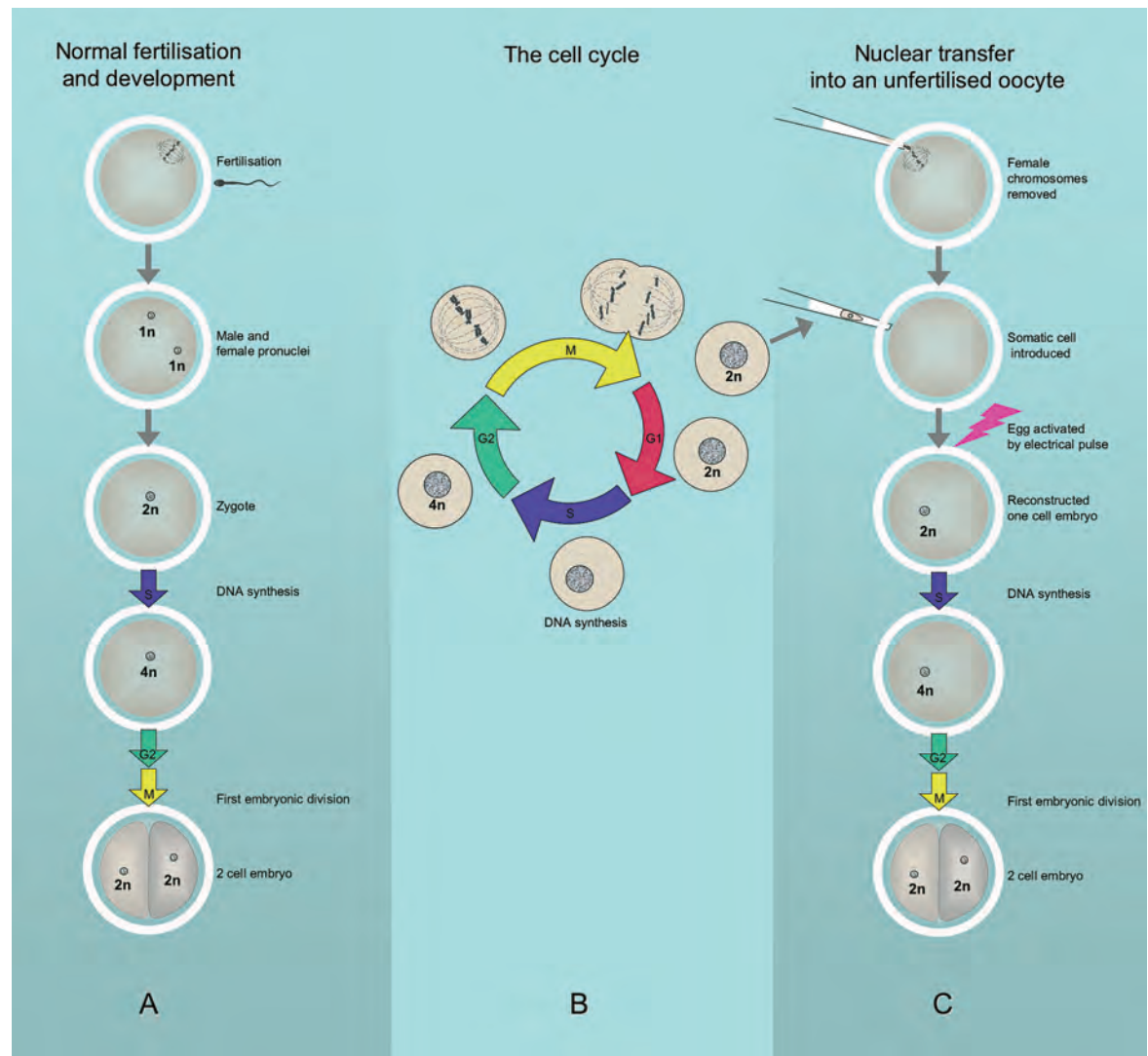


Figura 3. Transferencia nuclear y ciclo celular. A) Fertilización y desarrollo normal en el embrión en fase bicelular. B) El ciclo celular. A la fase G1 le siguen la fase S, cuando la célula duplica cada cromosoma, después la fase G2 y la mitosis (M), en la que el núcleo se rompe, los cromosomas duplicados se condensan, se alinean en el huso y se distribuyen en dos nuevas células hija. C) Transferencia nuclear utilizando una célula donante en la fase G1 y desarrollo hasta llegar al embrión en fase bicelular. 1n, 2n, 4n = copias de cada cromosoma.

del ciclo celular del donante nuclear afectaban al resultado de la transferencia nuclear. Fue una aportación determinante para el consiguiente éxito de la transferencia nuclear en mamíferos. Destacamos aquí los principales puntos.

Los ovocitos mamíferos se forman a partir de células germinales mediante un proceso llamado meiosis, que deja a cada ovocito sólo con una copia de cada cromosoma, cuya abreviatura suele ser «1n». Cuando la cabeza del espermatozoide entra, provoca que el ovocito termine la meiosis e inicie el desarrollo, tal y como ilustra la figura 3A. Los dos agrupamientos de cromosomas masculino y femenino se forman primero en pronúcleos separados, y después se unen para crear un único núcleo, en lo que es ahora el embrión unicelular o cigoto. Después, todos los cromosomas se replican mediante la síntesis del ADN preparada para la primera división de células embrionarias.

Esta primera división y todas las posteriores divisiones de células que forman y mantienen el cuerpo del animal se realizan mediante un proceso denominado mitosis. La mitosis es parte de un ciclo que garantiza que las células divididas conservan el número correcto de cromosomas. Este «ciclo celular» se divide de forma convencional en cuatro fases, tal y como se indica en la figura 3B. La primera fase se denomina gap1 (G1) en la que la célula tiene dos copias de cada cromosoma (2n). En la fase siguiente, la síntesis (S), la célula replica todo su ADN. Después viene la fase gap2 (G2) en la que cada cromosoma se presenta en cuatro copias (4n). En la mitosis (M), el núcleo se escinde, los cromosomas duplicados se condensan, se alinean en una estructura llamada huso y después se separan en dos nuevas células hija. Cada nueva célula contiene 2n cromosomas y el proceso se repite. Al dividirse rápidamente, el ciclo no lleva más de un día.

Esto tiene profundas implicaciones en la transferencia nuclear. Cuando un núcleo celular se transfiere a un ovocito no fertilizado tiene que ser activado, por ejemplo, por pulsación eléctrica, para desencadenar el inicio de su desarrollo. Esto pone en marcha la síntesis del ADN lista para la primera división celular. Sin embargo, esto ocurre con independencia de la fase del ciclo celular del núcleo donante. Si el núcleo entrante estaba en la fase S o G2 cuando el ADN ya se había replicado de forma parcial o total, su ADN se volverá a replicar, lo que provocará un número equivocadamente alto de cromosomas, o un daño cromosómico grave. El descubrimiento clave de Campbell y Wilmut fue que sólo las células donantes en fase G1 (antes de la replicación del ADN) soportarían un desarrollo normal en ovocitos no fertilizados, tal y como se muestra en la figura 3C.

El método que desarrollaron, y que se sigue utilizando, consiste en privar a las células donantes de factores de crecimiento reduciendo la cantidad de suero en el medio de cultivo durante unos días. Esto bloquea el ciclo celular antes de la fase S, que es exactamente lo que se quiere. Es importante destacar que la privación efectiva de suero exige que las células se cultiven durante varios días.

En 1995 nacieron Megan y Morag en el Roslin Institute, dos corderas creadas por transferencia de los núcleos a partir de células de un embrión de oveja de nueve días cultivadas por Jim McWhir y sometidas a entre 3 y 16 pases (Campbell et al. 1996). Estas ovejas no tardaron en suscitar un debate entre los coautores sobre cuál era el factor que propició el éxito del experimento. Según Campbell y Wilmut fue que la privación de suero antes de la transferencia nuclear no sólo coordinó el ciclo celular, sino que también indujo a un estado inactivo en el núcleo que lo hizo especialmente susceptible para la reprogramación por parte del ovocito. McWhir sostuvo que la clave residía en algunas propiedades de las células que había producido.

La transferencia nuclear en la oveja está dictada por la época natural de reproducción así que la pregunta no podía resolverse hasta el año siguiente. El plan original de 1996 consistía en volver a utilizar células embrionarias y también comprobar si la técnica podía ampliarse a células en una fase de desarrollo más avanzado, fibroblastos de un feto de 26 días. En ese momento estábamos trabajando con PPL Therapeutics, una compañía de biotecnología dedicada a producir proteínas farmacéuticas en la leche de ovejas transgénicas, y que estaba muy cerca del Roslin Institute. En el curso de una conversación durante una comida sugerimos un experimento más audaz y propusimos incluir células adultas en la época de transferencia nuclear de 1996. Esta idea fue acogida con escepticismo. Se dijo que era prematuro y que de todas formas no había fondos suficientes para ampliar el experimento. Sin embargo, si el trabajo adicional podía justificarse en términos comerciales, era posible que la compañía para la que trabajáramos aportara los fondos necesarios. Y así fue. Nos pusimos a investigar introduciendo transgenes de leche en células epiteliales mamarias de oveja que indujo Colin Wilde, del Hannah Research Institute de Ayr, como un medio para probar su expresión. Combinar ambos proyectos constituía una oportunidad idónea. Si las células mamarias podían convertirse en animales vivos, PPL tendría el potencial de producir «rebaños instantáneos» de ovejas que se sabía expresaban un transgén particular. Y, lo que es más interesante, utilizar células adultas para la transferencia nuclear abordaría la cuestión de la determinación celular, pendiente desde hacía mucho tiempo. Se presentó el caso a los directores de gestión e investigación de PPL, Ron James y Alan Colman, y la compañía se arriesgó a financiar el experimento. En febrero de 1996, se privó de suero a cultivos de células mamarias de oveja y también a células embrionarias cultivadas y se transportaron al Roslin Institute. Después, Bill Ritchie, el experto técnico de Wilmut, los transfirió a ovocitos enucleados de una oveja de raza Blackface (de cara negra) escocesa.

El 5 de julio nació una única cordera a la que su cuidador, John Bracken, llamó Dolly, en homenaje a Dolly Parton y a su gran talento como cantante. También nacieron dos corderas a partir de los fibroblastos fetales y cuatro de células

embrionarias. De esta forma se demostró que el éxito del experimento no se debía a ningún tipo de célula especial; la idea de que la inactividad había desempeñado también un papel se descartó posteriormente. Lo que se puso de manifiesto fue la importancia de la sincronización celular.

El 27 de febrero de 1997 se publicó una descripción del experimento (Wilmut et al. 1997). Más de una década después, podemos afirmar con tranquilidad que la transferencia nuclear a partir de una célula adulta descartó el concepto de determinación celular irreversible. Sin embargo, esto significaría pasar por alto la polémica que se prolongó hasta 17 meses después de su publicación. La idea de la determinación celular estaba tan asentada que varios prestigiosos científicos en Estados Unidos y Europa rechazaron el trabajo por considerarlo un fraude, recordando quizás la polémica de Illmensee. Un artículo del *New York Times* del 29 de julio de 1997 nos da una idea de la situación: «¿Cómo sabemos que todo eso no es más que un engaño? ¿Por qué, preguntan algunos, el resto del mundo se muestra tan dispuesto a aceptar la escandalosa noticia de que se ha clonado un animal adulto?».

Otros analistas defendieron la ineficacia de la transferencia nuclear adulta arguyendo que Dolly era una excepción, una aberración experimental. Muchos científicos eminentes sugirieron que no se había clonado a partir de una célula adulta, sino de material contaminante embrionario o fetal. Alguien dijo que las células fetales presentes en la circulación sanguínea de la oveja utilizadas para proporcionar las células mamarias habían penetrado de alguna manera los cultivos de células mamarias. Parecía que cualquier explicación alternativa, por poco plausible que resultara, era preferible a acabar con la doctrina de la determinación celular. La revista *Time* publicó un artículo el 2 de marzo de 1988 con el siguiente título «¿Fue Dolly un error?» que concluía diciendo: «En otras palabras, Dolly puede ser una casualidad y no un fraude. Pero sea lo que sea, cada vez es más probable que dentro de poco veamos clones de Bill Gates o Michael Jordan».

Mientras tanto, nosotros, y otros más, ya habíamos advertido de la existencia de más animales clonados (Schnieke et al. 1997), pero que se habían originado a partir de células fetales cultivadas y por tanto no confirmaban la clonación adulta.

Afortunadamente, las acusaciones y las especulaciones acabaron el 23 de julio de 1998. La edición de ese día de *Nature* publicaba dos destacados artículos. Uno ofrecía los resultados de un análisis independiente de huellas dactilares de ADN que confirmaban que el ADN nuclear de Dolly era idéntico a las células mamarias cultivadas (Signer et al. 1998). El segundo era un informe de Ryuzo Yanagimachi y Teruhiko Wakayama, de la Universidad de Hawai, en el que se describía otro animal clonado a partir de células adultas, un ratón al que se llamó «Cumulina» por las células del *cumulus* (folículo ovárico) utilizadas como donantes (Wakayama et al. 1998). La realidad de la clonación

adulta se aceptaba por fin. Ha pasado más de una década y es tiempo de pasar revista a los experimentos que se han realizado desde entonces.

Clonación con fines reproductivos

Inevitablemente, gran parte de los debates públicos, políticos y éticos se han centrado en la clonación para la reproducción de seres humanos y el resultado es que se promulgan en todo el mundo nuevas leyes y normativas. Es conveniente señalar que, en contra de lo que afirman algunos grupos religiosos y otros que buscan publicidad, ninguno de los científicos que trabajan en este campo consideró jamás llevar a cabo una clonación reproductiva en humanos.

MAMÍFEROS CLONADOS

Ganado, venado, gato, perro, hurón, cabra, gaur, caballo, ratón, muflón europeo, mula, cerdo, conejo, rata, mono rhesus, oveja, búfalo de la india, gato montés, lobo.

Como suele ser el caso, una vez que el método está bien establecido, es difícil ver por qué una vez se consideró imposible. Se ha empleado con éxito una variedad de diferentes tipos de células, tanto fetales como adultas, como donantes nucleares y se han clonado más de 20 especies, incluyendo peces, ranas, moscas de la fruta y los mamíferos que recoge la tabla. Sin embargo, la eficiencia es bastante reducida en la mayoría de las especies, ya que sólo entre el 1 y el 5% de los embriones reconstruidos llegaron a nacer. Los animales nacidos por transferencia nuclear pueden sufrir enfermedades, pero sus crías no, como es el caso de Bonny, el cordero de Dolly.

La mayoría de los experimentos de transferencia nuclear se siguen realizando en ganado. Los avances obtenidos son graduales más que espectaculares, aunque han alcanzado índices de éxito de alrededor del 15% en ganado. Un factor a destacar son los progresos en las técnicas de maduración de los ovocitos. Todos los ovocitos bovinos se obtienen ahora a partir de ovarios que se recogen en los mataderos en vez de extraerse del tracto reproductivo de animales vivos. Se obtiene así un suministro abundante de ovocitos que reduce en gran medida el número de animales que se necesitan. También permite la viabilidad comercial de la transferencia nuclear independientemente de que funcione o no, sobre todo en la reproducción de animales de élite con características muy atractivas, como es el caso de los caballos de carreras o toros para concursos. En Estados Unidos, compañías como ViaGen ofrecen clonación de ganado como parte de sus servicios de reproducción asistida. Su página web (www.viagen.com) dice: «ViaGen permite a los propietarios de ganado, caballos y cerdos conservar y multiplicar las mejores genéticas gracias a bancos de genes y servicios de clonación, y proteger sus marcas mediante servicios genómicos». A los leales (y ricos) propietarios de perros también les pue-

de interesar saber que «llamando al número gratuito 888-8VIAGEN pueden conocer las opciones disponibles para clonar su perro».

La transferencia nuclear se ha utilizado cuando la reproducción sexual normal es imposible como resultado de un accidente, enfermedad o infertilidad natural, tal y como se demostró clonando una mula (Woods et al. 2003). También se ha aplicado para reproducir especies en peligro de extinción como el gato montés europeo o razas de ganado poco frecuentes. Sin embargo, sólo aquellas especies con parientes domesticados que proporcionen ovocitos adecuados podrán beneficiarse de esta técnica. La clonación tampoco sirve para mejorar la diversidad genética de poblaciones animales reducidas, que es vital para la supervivencia a largo plazo.

Animales para la biomedicina

El futuro nos dirá si la transferencia nuclear puede llegar a ser otra técnica habitual, aunque costosa, de reproducción animal. Pero lo cierto es que ya se ha convertido en uno de los mejores métodos para instaurar linajes de grandes animales «transgénicos», es decir, modificados genéticamente. Para producir animales transgénicos existen dos opciones. El ADN transgénico puede introducirse directamente dentro del cigoto, con la esperanza de que se incorpore al genoma. Alternativamente, las modificaciones genéticas pueden realizarse en células cultivadas y utilizarse después para producir animales completos. El primer método es una suerte de lotería, ya que los animales deben producirse antes de analizar si está presente un transgén. El segundo método permite ejercer mucho más control e implica a un número menor de animales porque las células pueden analizarse minuciosamente en el laboratorio antes de que se produzca ningún animal.

En ratones, un método basado en la célula está disponible desde principios de la década de 1980. Las células madre embrionarias (en inglés ES) de ratones pueden aislarse en embriones tempranos, crecer indefinidamente mediante cultivo, ser objeto de manipulaciones como la incorporación de un transgén o la alteración de un gen en particular (genes diana), y luego pueden volver a incorporarse al embrión en desarrollo. El enorme potencial de la tecnología de genes diana en las células ES nos ha proporcionado casi todos los conocimientos de que disponemos sobre la función de los genes en animales completos. Así lo reconoció el Premio Nobel de Medicina de 2007, otorgado conjuntamente a Mario Capecchi, Martin Evans y Oliver Smithies. Hace tiempo que muchos investigadores saben que ampliar esta técnica a animales grandes tendría muchas y útiles aplicaciones. Pero a pesar de los intentos, las células ES funcionales no han podido y siguen sin poder obtenerse del ganado. La transferencia nuclear en ganado utilizando células somáticas que podrían prosperar y someterse a manipulación en cultivo superó claramente esta técnica.

En los experimentos que siguieron a Dolly demostramos que tanto las ovejas transgénicas y con genes diana pueden generarse mediante transferencia nuclear (Schnieke et al. 1997, McCreath et al. 2000). Desde entonces se han realizado muchos otros, por ejemplo, ganado con genes diana resistente a la enfermedad de las vacas locas (Kuroiwa et al. 2004). Sin embargo, gran parte de las aplicaciones se han realizado en el campo de la biomedicina, y los animales transgénicos que producen proteínas farmacéuticas en la leche han llegado más tarde de lo que se esperaba. ATryn, un fármaco anticoagulante utilizado para tratar pacientes con una deficiencia hereditaria de antitrombina, la proteína regula la coagulación de la sangre, se ha producido en cabras transgénicas a partir de un primer animal clonado, y GTC Biotherapeutics lo comercializó en noviembre de 2007. La transferencia nuclear también se utiliza para realizar modificaciones genéticas múltiples en cerdos con el fin de producir células u órganos que puedan luego trasplantarse a seres humanos, una técnica denominada xenotransplante.

También se han desarrollado una serie de modelos de grandes animales de enfermedades humanas graves como la fibrosis quística (Rogers et al. 2008). Suelen ser ampliaciones de trabajos llevados a cabo en ratones, donde la técnica de genes diana ha proporcionado una considerable información en relación con enfermedades como el cáncer. Se han producido muchas variedades de ratones con defectos genéticos, y han sido muy valiosas para comprender los mecanismos del inicio y el avance de los tumores (Frese y Tuveson 2007). Los ratones también son útiles en los estudios de viabilidad de las pruebas en diagnósticos nuevos y estrategias de tratamiento, de lo que hablaremos más adelante. Sin embargo, las importantes diferencias en tamaño del cuerpo, fisiología general, anatomía, dieta y esperanza de vida limitan la utilidad de los ratones. Por ejemplo, la radiación y la terapia térmica no pueden reducir su escala para tratar tumores de ratones. La transferencia nuclear brinda la oportunidad de ampliar la gama de modelos de enfermedades definidas genéticamente a otras especies, como cerdos, que se parecen más a los seres humanos en lo que se refiere a tamaño, anatomía y fisiología.

Transferencia nuclear, células madre embrionarias y medicina regenerativa

Tal y como se ha mencionado anteriormente, gran parte del interés de la transferencia nuclear a finales de la década de 1980 y principios de la de 1990 se debió a las posibilidades que ofrecía la tecnología de células madre embrionaria. Desde entonces, ambos campos se han entrelazado íntimamente.

Las células ES suelen aislarse a partir de embriones en fase de blastocisto. Un blastocisto es una diminuta bola llena de fluido de un centenar de células contenidas dentro de un racimo de células denominada masa celular interna (ICM) que da origen a todos los tejidos del cuerpo.

Los blastocistos, o ICM aislados, se cultivan y en unos días o una semana emergen colonias de pequeñas células muy concentradas que además siguen creciendo indefinidamente; son células ES. Por motivos que no se conocen con certeza, obtener células ES resulta bastante difícil en el caso de muchas especies y sólo ha sido posible en ratones, seres humanos y monos Rhesus. Se ha informado a la prensa de la obtención de células ES en ratas, pero todavía no se han publicado estos hallazgos en una revista científica.

Las células ES suelen utilizarse como un sucedáneo apropiado para el estudio del embrión temprano, pero sigue sin estar claro lo que son en realidad. Pueden ser un artefacto de cultivo de tejidos, algo aberrante creado como respuesta a las condiciones de crecimiento artificial. Sin embargo, pruebas recientes sugieren que son un tipo de célula presente durante un periodo corto de tiempo en el embrión, que puede ser capturado y conservado con unas condiciones de cultivo adecuadas (Silva y Smith 2008).

La característica definitoria de las células madre embrionarias es que pueden crecer de forma indefinida como células no diferenciadas y luego diferenciarse en muchos otros tipos de células. Cuando se introducen en un blastocisto pueden integrarse en el ICM y participar en la formación de todos los tejidos del cuerpo. Cuando se aplican los estímulos apropiados también pueden formar una amplia variedad de tipos de células en cultivo, que se denomina diferenciación *in vitro*. Desde que Jamie Thomson obtuvo por primera vez células ES humanas hace diez años, (Thomson et al. 1998) ha surgido un gran interés por la diferenciación *in vitro* como una posible fuente de sustitución de tejido humano, como las células nerviosas, células que producen insulina, o células musculares del corazón. Se han escrito muchos artículos sobre este tema, así que no entraremos en más detalles. El esquema básico se muestra en el panel A de la figura 4. La promesa de la terapia basada en ES es ya una realidad, pero puede que sean necesarias algunas palabras de cautela. Conseguir que las células madre embrionarias humanas generen cantidades útiles de tipos de células adecuadas, que estén apropiadamente caracterizadas para fines terapéuticos supone es un gran reto. Además, es necesario establecer métodos rigurosos que garanticen que los preparados obtenidos a partir de ES están libres de células que podrían formar tumores. Los científicos investigadores y la industria biotecnológica deberían ser realistas y evitar esa tendencia a airear a bombo y platillo sus descubrimientos.

Seguramente, Geron, la compañía farmacéutica de California, es la que más ha avanzado en terapias celulares ES humanas. Su página web (www.geron.com) informa sobre el desarrollo de células progenitoras nerviosas humanas obtenidas de células ES para lesiones graves de la columna vertebral y cardiomiocitos para el tratamiento de fallos cardíacos. Geron ha solicitado permiso para llevar a cabo pruebas clínicas con células progenitoras nerviosas en seres humanos, pero la United States Food and

Drug Administration no ha dado todavía su aprobación. Si se autorizan estos experimentos, el resultado será determinante para terapias futuras con células ES.

Si pueden producirse, los tejidos obtenidos de células ES tendrían que coincidir inmunológicamente con el paciente de la misma forma que un tejido donante normal, para evitar el rechazo. Además, es muy posible que los receptores precisen supresión inmunológica durante toda su vida. La coincidencia de los tejidos resulta problemática en pacientes con tipos de tejidos poco comunes, como las personas de raza mixta. Poco después del informe de Thomson se sugirió que la transferencia nuclear podría proporcionar un método para producir tejidos humanos a medida mediante «clonación terapéutica». Las células podrían extraerse de un paciente humano que necesitase una terapia de sustitución de tejidos y utilizarse para producir embriones clonados. Se obtendrían células ES y después se inducirían para su diferenciación en cultivo. El tejido de sustitución coincidiría perfectamente con el propio cuerpo del paciente (fig. 4B).

Se ha logrado dar algunos de los pasos necesarios con animales. Por ejemplo, se han obtenido células ES a partir de embriones de ratones por transferencia nuclear y se ha descubierto que son las mismas que las obtenidas de embriones normales. También se han producido células ES de monos Rhesus a partir de embriones clonados, pero hasta el momento no se han obtenido células madre embrionarias humanas por transferencia nuclear. El mayor obstáculo práctico es el suministro de ovocitos humanos no fertilizados, que ya es insuficiente para satisfacer las necesidades de la gente que desea someterse a las técnicas de reproducción asistida como la fertilización *in vitro* (FIV). Un reciente artículo en *Nature* revelaba que, a pesar de que han pasado dos años y se han invertido 100.000 dólares en publicidad local, los investigadores de células madre de la Universidad de Harvard sólo han conseguido una donante de óvulo (Maher 2008).

Esto quiere decir que es muy improbable que la clonación terapéutica se convierta en una realidad hasta que no se encuentre una fuente alternativa de ovocitos receptores. Hay muchos ovocitos animales, sobre todo de ganado, gracias a la maduración *in vitro*. La Human Fertilization and Embryology Authority (HFEA) del Reino Unido aprobó recientemente su investigación, pero son muchos quienes se oponen a la creación de embriones híbridos citoplásmicos. Los problemas biológicos también pueden surgir de la incompatibilidad entre los componentes del ovocito animal y del núcleo humano entrante. Los factores de reprogramación e importantes organelos celulares como las mitocondrias pueden no funcionar correctamente. Puede que la fuente más prometedora de ovocitos sea la maduración *in vitro* de ovocitos humanos inmaduros provenientes de ovarios donados. A pesar de que esté menos avanzada que en el ganado, la maduración *in vitro* de ovocitos humanos está mejorando, y se utiliza sobre todo

para ayudar a las mujeres que han tenido que someterse a una ovariectomía. Se tiene conocimiento de varios nacimientos normales a partir de ovocitos madurados *in vitro*.

A pesar de sus posibles beneficios, la derivación de células madre embrionarias humanas y la clonación terapéutica se enfrentan a una fuerte oposición religiosa y ética, y está claro que mucha gente no aceptará la generación artificial ni la destrucción de embriones humanos, a pesar de que muchos consideran que sólo son diminutas bolas de células. Las leyes son diferentes en todo el mundo. Por ejemplo, la HFEA inglesa dio su aprobación en 2007, el Ministerio de Sanidad español aprobó la investigación en 2008, mientras que en Alemania no existen planes para legalizar este tipo de procedimientos. Los ele-

vados costes y el tiempo requerido suponen un problema añadido. Lo más probable es que la clonación terapéutica quede limitada a pacientes ricos y sólo si su estado de salud les permite esperar varios meses. Dicho esto, los recientes avances en reprogramación nuclear han dejado obsoleta la reprogramación terapéutica.

Comprender la reprogramación

Un cuerpo humano contiene varios cientos de tipos de células, y cada uno de ellos difiere en una multitud de componentes celulares. La identidad de una célula, su forma, la rapidez con la que se divide, los materiales que sintetiza, los receptores que están en su superficie, y las múltiples nanomáquinas que llamamos ARN (ácido ribo-

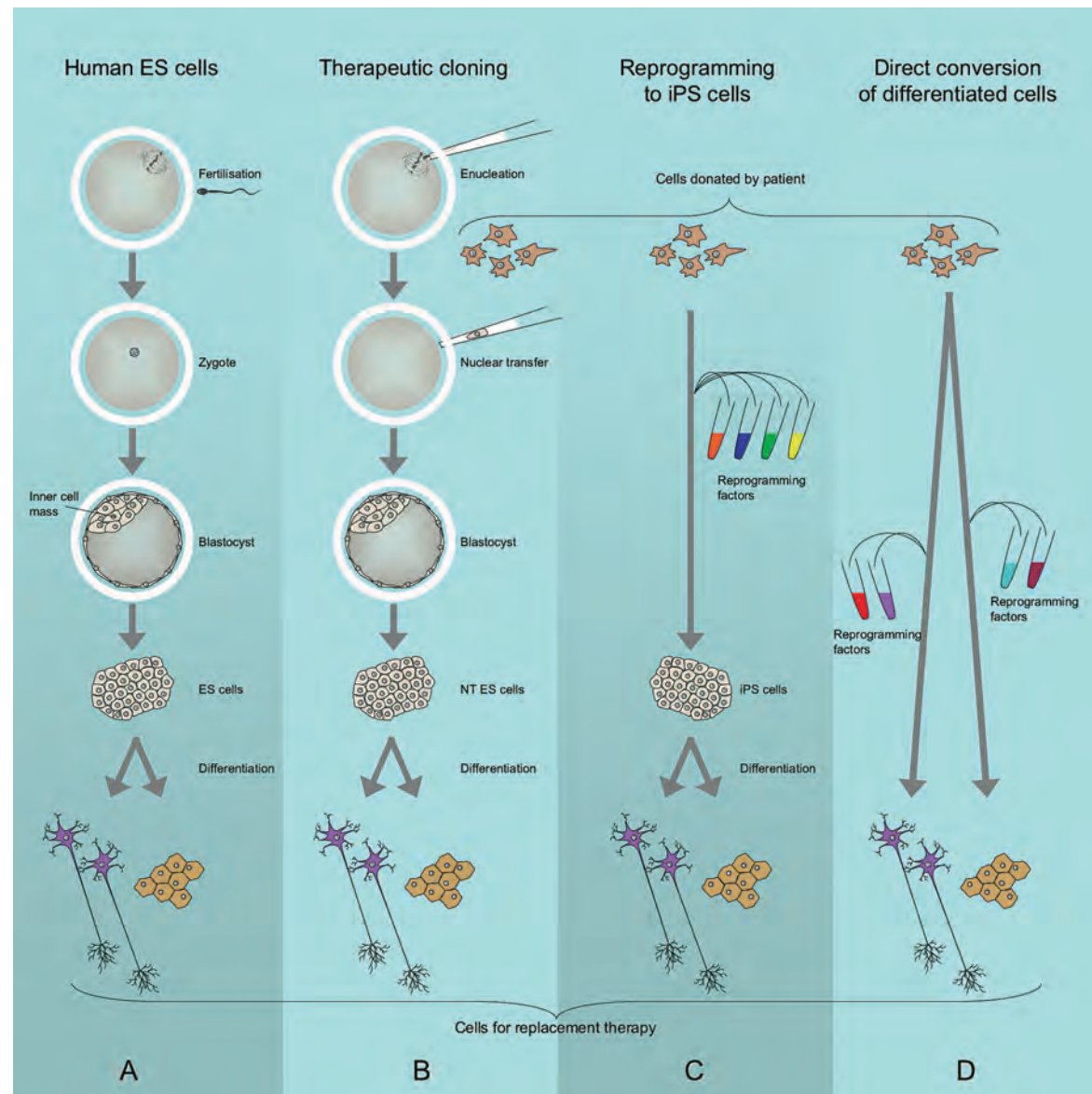


Figura 4. Generación de células diferenciadas para la investigación de enfermedades y la terapia celular. A) Derivación y diferenciación celular ES humana estándar. B–D) Derivación de células específicas de pacientes. B) Clonación terapéutica. C) Células iPS. D) Conversión directa de un tipo de célula diferenciada a otro. ES, células madre embrionarias; NT-ES, células ES derivadas por transferencia nuclear; iPS, células madre pluripotentes inducidas.

nucleico) y moléculas proteicas, son todos ellos el producto de diferentes modelos de expresión genética.

Clonar células adultas demostró que estos modelos no se deben a diferencias genéticas inmutables. Los núcleos de las células más diferenciadas, como las neuronas o los linfocitos B maduros especializados en sintetizar un único anticuerpo, conservan el potencial para formar todas las células del cuerpo (Hochedlinger y Jaenisch 2002). Cada célula tiene la misma información genética; en el caso de los seres humanos unos 3.000 millones de pares de bases de ADN y una cifra aproximada de 25.000 genes, pero éstos se expresan de forma diferente. Se puede trazar un paralelismo con el *software* y el *hardware* informático. Hay diferentes programas para los gráficos, para las matemáticas, para la música o para el tratamiento de textos, pero todos pueden funcionar en una misma máquina sin alterar sus componentes físicos. Por supuesto, a diferencia de los ordenadores, las células se organizan por sí solas y resulta inimaginable pensar que una célula pueda recibir un juego completo de instrucciones por parte de lo que llamaríamos un agente externo.

La regulación de la expresión genética lleva estudiándose más de 30 años, y se sabe que funciona a varios niveles: desde la accesibilidad del ADN dentro del núcleo hasta los factores de expresión; el índice al que los genes se transcriben en moléculas RNA mensajeras; el procesamiento y transporte del ARN, hasta la síntesis y degradación de los productos proteínicos. Si echamos un vistazo a los diagramas de un libro de texto moderno de biología celular o molecular, nos encontramos con multitud de flechas que representan los caminos regulatorios y los bucles de *feedback* que rigen el funcionamiento de nuestras células. Se suelen utilizar los términos «circuitos moleculares» o «redes de genes». La identidad de una célula es el producto de un complejo entramado de interacciones y tiene un estado dinámico, no estático. Se sabe que los factores regulatorios van constantemente del núcleo al citoplasma, afectando a su propia expresión y a la de otros genes. Se puede decir por tanto que una célula está constantemente actualizando su programa de expresión genética. Si el ciclo regulatorio favorece la continuación de un estado concreto, éste es estable. Y de la misma forma que una red física puede adoptar diferentes formas cuando se tira de ella o se la empuja, existen muchos modelos estables de expresión genética y muchos estados diferenciados.

En una célula en particular, algunos genes se expresan en gran medida, otros menos y otros nada. No se conoce por completo el proceso mediante el cual el modelo se mantiene y se transmite de forma fiable a las células hija. Pero sí se sabe que la cromatina, el complejo de ADN que rodea las proteínas histonas, lleva marcas que informan si los genes están activos o inactivos. Estas marcas son «epigenéticas» más que genéticas, en el sentido que no alteran la verdadera secuencia de ADN. Por ejemplo, el ADN que se encuentra dentro y alrededor de genes inactivos transporta

a menudo grupos de metilo añadidos a la citosina nucleotídica. Los genes activos e inactivos también muestran modificaciones químicas diferentes ante los histones. Esto afecta a lo fuerte que está encadenado el ADN y a lo «abierto» y receptivo que esté a los factores de transcripción.

Cuando un núcleo de una célula diferenciada se expone a un entorno extraño, por ejemplo cuando dos células se fusionan, los procesos regulatorios se interrumpen y el modelo de expresión genética se ve alterado de la misma forma. Por ejemplo, un núcleo de un hígado humano puede ser inducido para expresar genes musculares con una célula muscular de ratón (Blau, Chiu y Webster 1983) y los núcleos de varias células somáticas expresan genes embrionarios cuando se fusionan con células ES (Do, Han y Schöler 2006).

Se puede decir que la transferencia genética es una versión más completa del mismo fenómeno. Cuando se transfiere un núcleo a un ovocito enucleado, sufre una eliminación completa de los grupos de metilo de ADN y cambios significativos en modificaciones de histonas, con lo que se borra por completo su identidad previa. Kevin Eggan y sus colegas afirman que la clave del éxito de dicha reprogramación reside en la libre disponibilidad de los factores que regulan la transcripción genética (Egli, Birkhoff y Eggan 2008). Éstos se asocian normalmente con el ADN dentro del núcleo, pero se liberan en el citoplasma cuando el núcleo se escinde y se prepara para repartirse con los cromosomas en los dos nuevos núcleos. Los ovocitos sin fertilizar tienen gran abundancia de dichos factores libres, y están sanos y preparados para reprogramar un núcleo entrante; la cabeza del espermatozoide.

Pero ¿qué factores son responsables de reprogramar un núcleo a un estado embrionario? Por desgracia, los ovocitos de los mamíferos son diminutos, no se propagan, y por lo tanto son difíciles de analizar con la tecnología de la que se dispone actualmente. Por esa razón los investigadores se han centrado en las células madre embrionarias ES.

Reprogramación directa, un enfoque radicalmente nuevo

Años de intenso trabajo han revelado mucho sobre los mecanismos que mantienen las células ES en un estado indiferenciado y que desencadenan su diferenciación. En 2006 todo ello culminó en un importante hallazgo. Shinya Yamanaka y sus colegas de la Universidad de Kioto llegaron a la conclusión de que los factores regulatorios que se sabe son importantes para mantener las células ES indiferenciadas serían buenos candidatos para reprogramar los factores. Su grupo identificó 23 genes regulatorios y construyó vectores virales para transducirlos individualmente en otras células. Después se introdujeron varias combinaciones de genes en los fibroblastos de ratón y se seleccionaron células para la expresión de un gen expresado de forma característica en células ES. Se encontró un juego de cuatro factores de transcripción: Sox-2, Oct-4, c-Myc y Klf4, para convertir los fibroblastos en algo que se

pareciese a las células ES, que denominaron células madre (iPS) pluripotentes inducidas (Takahashi y Yamanaka 2006). Después de este primer estudio, Yamanaka y otros grupos de investigadores han refinado la técnica y la han ampliado a las células humanas (fig. 4C). En el momento de escribir estas líneas, la opinión general es que, en esencia, las células iPS y las células ES son las mismas. Sin embargo, todavía es pronto para realizar afirmaciones y algunos críticos han señalado que las diferencias pueden ser significativas (Liu 2008).

El descubrimiento de una receta tan increíblemente sencilla para reprogramar células diferenciadas a un estado embrionario ha desatado una actividad investigadora frenética en todo el mundo. A diferencia de la transferencia nuclear, no hay problemas éticos y las técnicas son fáciles, lo que abre el estudio de la reprogramación a muchos laboratorios. Además, algunos destacados grupos que se dedicaban a la investigación de la clonación terapéutica han cambiado de investigación. El *US Boston Globe* del 1 de agosto de 2008 citaba a Rudolf Jaenisch afirmando que el enfoque iPS «es mucho más fácil, y tiene muchas menos limitaciones y problemas, tanto éticos como de otra índole [...] Creo que avanzaremos en esta dirección».

El estudio de las células IPS se mueve a tal velocidad que es probable que este artículo esté ya obsoleto cuando se publique. Pero a medida que avanzan los estudios, se realizan importantes hallazgos. Al principio parecía que algunas células podían reprogramarse y otras no. Pero las células iPS se hacen ahora a partir de muchos tipos de células, como los linfocitos B maduros y las células beta de islotes pancreáticos, lo que demuestra que no se trata de un tipo de célula particularmente raro, ni de un artefacto experimental, como habían afirmado algunos escépticos. Aunque resulte desconcertante, diversos grupos de investigación están descubriendo que en general, las combinaciones diferentes y los números de factores son efectivas. Los mecanismos subyacentes siguen siendo un misterio, pero no por mucho tiempo. El análisis bioinformático de constelaciones completas de genes está revelando los modelos de expresión y los datos de las redes regulatorias que caracterizan a las células ES e iPS y a los acontecimientos implicados en la reprogramación directa (Mikkelsen et al. 2008; Müller et al. 2008).

Una aplicación inmediata de las células iPS es el estudio de enfermedades degenerativas. Las células iPS humanas ya se han aislado en pacientes con enfermedades de las neuronas motrices, la enfermedad del Parkinson, la distrofia muscular de Duchenne y la diabetes juvenil (tipo I) (Dimos et al. 2008; Park et al. 2008), y se están utilizando para generar placas del tipo de célula afectada en el laboratorio. La disponibilidad de las células iPS para enfermedades específicas tendrá un profundo impacto en la comprensión y tratamiento de muchos desórdenes graves. Permitirá que el efecto de los factores ambientales como los aditivos alimenticios, las toxinas o los patógenos en la degeneración

celular puedan ser examinados exhaustivamente en estudios a gran escala. También se podrán analizar muchos medicamentos con el fin de identificar cuáles pueden detener, ralentizar o revertir el avance de la enfermedad.

Las células IPS también han despertado un gran interés como fuente de sustitución de tejidos sin la necesidad de contar con óvulos humanos o células ES. En un experimento de prueba de concepto, Rudolf Jaenisch trató con éxito la anemia drepanocítica en ratones (Hanna et al. 2007). Las células de piel de un ratón con anemia drepanocítica se convirtieron en células iPS y el defecto genético se corrigió mediante genes diana. Las células iPS se indujeron para diferenciarse en células madre sanguíneas y después se trasplantaron al ratón, donde redujeron la anemia y aumentaron sus posibilidades de supervivencia.

Sin embargo, es preciso subrayar que dichas terapias basadas en células iPS están todavía lejos de las aplicaciones clínicas humanas. Los métodos actuales de producir células iPS incluyen algunos genes cancerígenos peligrosos, así que habrá que encontrar otras alternativas. Pero resulta esperanzador que ya existan algunos estudios previos que indican que los agentes químicos pueden sustituir la necesidad de algunos genes en la receta original, y variaciones como añadir ARN instructivo en vez de genes también se están estudiando.

Si las células iPS son idénticas a las células ES, se enfrentan necesariamente a los mismos problemas en lo que se refiere a su uso terapéutico. Al igual que las células ES, las células iPS no diferenciadas pueden formar tumores, por lo que deben quedar absolutamente excluidas de cualquier medicamento terapéutico. También se deben crear métodos para inducir la diferenciación de poblaciones puras de tipos de células terapéuticas. Se han establecido condiciones de diferenciación para algunos de ellos, como las neuronas motoras, pero se deben crear procedimientos para muchas otras células potencialmente útiles.

Los dos años de historia de la revolución iPS han sido asombrosos, y casi han dejado obsoleta la clonación terapéutica. Pero ahora hay signos de otro cambio más. Aunque el ovocito era una caja negra que no revelaba fácilmente sus funciones, la derivación de células iPS ha abierto las puertas al estudio de la reprogramación. Cada vez son más los conocimientos sobre el desarrollo normal de muchos tipos de células, y el papel de las moléculas regulatorias clave cada vez está más claro, lo que permite que se utilicen como «botones e interruptores» para controlar la identidad celular. Si el objetivo es producir células diferenciadas para ordenarlas, ¿por qué hacerlo directamente sin utilizar un intermediario embrionario? En una espectacular publicación (Zhou et al. 2008) de agosto de 2008, Doug Melton y sus colegas informaban de que habían tratado a ratones diabéticos con tres genes instructivos transportados en vectores virales. Introdujeron algunas de las células exocrinas en el páncreas de los ratones, que normalmente segregan enzimas digestivas,

para convertirlas directamente en insulina produciendo células beta sin ninguna formación intermediaria de IPS, u otras células de tipo ES. No cabe duda de que este trabajo se encuentra en su fase inicial, pero ha abierto otro camino a la producción de células (fig. 4D). Pero lo que resulta más provocador es que, al haberse realizado el estudio en ratones y no en cultivos, ahora hay pacientes que exigen una terapia de sustitución y que son capaces de ingerir un coctel de factores instructivos diseñados para generar nuevas células en su propio cuerpo sin necesidad de un trasplante. ¿Podría estar también en el horizonte la regeneración de órganos completos?

Conclusiones

En un reciente ensayo (Thomas 2007), John Meurig Thomas destacó la impredecibilidad básica del progreso científico y los tortuosos caminos que a menudo separan los

primeros descubrimientos de las investigaciones y el desarrollo de dispositivos y procedimientos modernos que nos son familiares. Es ya sabido que después de hacer público su invento, en 1958, Charles Townes y Arthur Schawlow no previeron una aplicación práctica para el láser óptico.

Originalmente, la clonación se concibió para investigar la determinación de la identidad celular y su destino, pero ahora está más orientada hacia la habilidad de cambiar el destino celular. ¿Quién sabe cuál acabará siendo el mayor legado de Dolly? Después de más de once años, es evidente que ha logrado que la gente se formule estas preguntas, y la sensación generalizada es que hay muchas puertas abiertas. Científicos de talento se han sentido atraídos por esta investigación y han emprendido proyectos que habrían sido inconcebibles antes de 1997. Creemos que el auge de la investigación actual aportará significativos avances en medicina y beneficiará a la salud humana.

Bibliografía

- Beetschen, J. C., y J. L. Fischer. «Yves Delage (1854–1920) as a forerunner of modern nuclear transfer experiments». *Int. J. Dev. Biol.* 48 (2004): 607–612.
- Blau, H. M., C. P. Chiu y C. Webster. «Cytoplasmic activation of human nuclear genes in stable heterocaryons». *Cell* 32 (1983): 1.171–1.180.
- Briggs, R., y T. J. King. «Transplantation of living nuclei from blastula cells into enucleated frogs' eggs». *Proc. Natl. Acad. Sci. EE.UU.* 38 (1952): 455–463.
- Campbell, K. H., J. McWhir, W. A. Ritchie y I. Wilmut. «Sheep cloned by nuclear transfer from a cultured cell line». *Nature* 380 (1996): 64–66.
- Dimos, J. T., K. T. Rodolfa, K. K. Niakan, L. M. Weisenthal, H. Mitsumoto, W. Chung, G. F. Croft, et al. «Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons». *Science* 321 (2008): 1.218–1.221.
- Do, J. T., D. W. Han, y H. R. Schöler. «Reprogramming somatic gene activity by fusion with pluripotent cells». *Stem Cell Rev.* 2 (2006): 257–264.
- Driesch H. «Entwicklungsmechanisme Studien. I. Der Werth der beiden ersten Furchungszellen in der Echinodermentwicklung. Experimentelle Erzeugen von Theil und Doppelbildung». *Zeit. für wiss. Zool* 53: 160–178, 1892, 183–184.
- Egli, D., G. Birkhoff y K. Eggan. «Mediators of reprogramming: transcription factors and transitions through mitosis». *Nat. Rev. Mol. Cell. Biol.* 9 (2008): 505–516.
- Frese, K. K. y D. A. Tuveson. «Maximizing mouse cancer models». *Nat. Rev. Cancer* 7 (2007): 645–658.
- Gurdon, J. B. «The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles». *J. Embryol. Exp. Morphol.* 10 (1962): 622–640.
- Hanna, J. M. Wernig, S. Markoulaki, C. W. Sun, A. Meissner, J. P. Cassady, C. Beard, et al. «Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin». *Science* 318 (2007): 1.920–1.923.
- Hochedlinger, K. y R. Jaenisch. «Monoclonal mice generated by nuclear transfer from mature B and T donor cells». *Nature* 415 (2002): 1.035–1.038.
- Illmensee, K. y P. C. Hoppe. «Nuclear transplantation in *Mus musculus*: developmental potential of nuclei from preimplantation embryos». *Cell* 23 (1981): 9–18.
- Kuroiwa, Y., P. Kasinathan, H. Matsushita, J. Sathiyaselan, E. J. Sullivan, M. Kakitani, K. Tomizuka, I. Ishida y J. M. Robl. «Sequential targeting of the genes encoding immunoglobulin- μ and prion protein in cattle». *Nat. Genet.* 36 (2004): 775–780.
- Liu, S. V. «iPS cells: a more critical review». *Stem Cells Dev.* 17, 2008: 391–397.
- Lysenko, T. D. *Soviet Biology: Report and concluding remarks to the 1948 session of the Lenin Academy of Agricultural Sciences*. (Edición inglesa) Londres: Birch Books, 1948. Versión online: www.marxists.org/reference/archive/lysenko/works/1940s/report.htm
- McCreath, K. J., J. Howcroft, K. H. S. Campbell, A. Colman, A. E. Schnieke y A. J. Kind. «Production of gene-targeted sheep by nuclear transfer from cultured somatic cells». *Nature* 405 (2000): 1.066–1.069.
- McGrant, J. y D. Solter. «Inability of mouse blastomere nuclei transferred to enucleated zygotes to support development in vitro». *Science* 226 (1984): 1.317–1.319.
- Maher, B. «Egg shortage hits race to clone human stem cells». *Nature* 453, 2008, 828–829.
- Mikkelsen, T. S., J. Hanna, X. Zhang, M. Ku, M. Wernig, P. Schorderet, B. E. Bernstein, R. Jaenisch, E. S. Lander y A. Meissner. «Dissecting direct reprogramming through integrative genomic analysis». *Nature* 454, 2008, 49–55.
- Morgan, T. H., A. H. Sturtevant, H. J. Muller y C. B. Bridges. *The Mechanism of Mendelian Heredity*. Nueva York: Henry Holt and Co., 1915.
- Müller, F. J., L. C. Laurent, D. Kostka, I. Ulitsky, R. Williams, C. Lu, I. H. Park, et al. «Regulatory networks define phenotypic classes of human stem cell lines». *Nature* 455 (2008): 401–405.
- Park, I. H., N. Arora, H. Huo, N. Maherali, T. Ahfeldt, A. Shimamura, M. W. Lensch, C. Cowan, K. Hochedlinger y G. Q. Daley. «Disease-specific induced pluripotent stem cells». *Cell* 134 (2007): 877–886.
- Rogers, C. S., Y. Hao, T. Rokhlina, M. Samuel, D. A. Stoltz, Y. Li, E. Petroff, et al. «Production of CFTR-null and CFTR-DeltaF508 heterozygous pigs by adeno-associated virus-mediated gene targeting and somatic cell nuclear transfer». *J. Clin. Invest.* 118 (2008): 1.571–1.577.
- Schnieke, A. E. A. J. Kind, W. A. Ritchie, K. Mycock, A. R. Scott, M. Ritchie, I. Wilmut, A. Colman y K. H. S. Campbell. «Human factor IX transgenic sheep produced by transfer of nuclei from transfected fetal fibroblasts». *Science* 278 (1997): 2.130–2.133.
- «Production of gene-targeted sheep by nuclear transfer from cultured somatic cells.» *Nature* 405, 1997b, 1066–1069.
- Signer, E. N., Y. E. Dubrova, A. J. Jeffreys, C. Wilde, L. M. Finch, M. Wells y M. Peaker. «DNA fingerprinting Dolly». *Nature* 394 (1998): 329–330.
- Silva, J. y A. Smith. «Capturing pluripotency». *Cell* 132 (2008): 532–536.
- Spemann, H. «Die Entwicklung seitlicher und dorso-ventraler Keimhälfen bei verzögert-er Kernversorgung». *Zeit. für wiss. Zool* 132 (1928): 105–134.
- , *Experimentelle Beiträge zu einer Theorie der Entwicklung*. Berlin: Springer, 1936. (Edición inglesa, *Embryonic development and induction*, 1938.)
- Takahashi, K. y S. Yamanaka. «Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors». *Cell* 126 (2006): 663–76.
- Thomas, J. M. «Unpredictability and chance in scientific progress». *Progress in Informatics* 4 (2007): 1–4.
- Thomson, J. A., J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall y J. M. Jones. «Embryonic stem cell lines derived from human blastocysts». *Science* 282 (1998): 1.145–1.147.
- Wakayama, T., A. C. Perry, M. Zuccotti, K. R. Johnson y R. Yanagimachi. «Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei». *Nature* 394 (1998): 369–374.
- Weismann, A. *Das Keimplasma. Eine Theorie der Vererbung*. Jena: Gustav Fischer, 1892.
- Willadsen, S. M. «Nuclear transplantation in sheep embryos». *Nature* 320 (1986): 63–65.
- Wilmut, I., A. E. Schnieke, J. McWhir, A. J. Kind y K. H. Campbell. «Viable offspring derived from fetal and adult mammalian cells». *Nature* 385 (1997): 810–813.
- Woods, G. L., K. L. White, D. K. Vanderwall, G. P. Li, K. I. Aston, T. D. Bunch, L. N. Meerdo y B. J. Pate. «A mule cloned from fetal cells by nuclear transfer». *Science* 301 (2003): 1.063.
- Zhou, Q., J. Brown, A. Kanarek, J. Rajagopal y D. A. Melton. «In vivo reprogramming of adult pancreatic exocrine cells to beta-cells». *Nature*, 27 de agosto 2008. [existe un formato ePub anterior]

hacia una comprensión del cáncer

JOAN MASSAGUÉ

Ha llegado la era en que la ciencia puede conquistar el cáncer y su reputación de enfermedad incurable empieza a desaparecer. Cada día nuevos avances en la investigación médica y la atención clínica contribuyen a desmontar el mito de que el cáncer es una enfermedad demasiado compleja de entender y difícil de curar. El camino a la comprensión y el control del cáncer sigue siendo arduo, pero los recientes progresos proporcionan razones para un optimismo cauto.

Verdades y mitos sobre el cáncer

El cáncer es una enfermedad en la que las células se multiplican de forma descontrolada, invaden los tejidos circundantes y se extienden a distintas partes del organismo en un proceso llamado metástasis. La invasión y la metástasis son los rasgos clave que distinguen los tumores malignos —es decir, el cáncer propiamente dicho— de los benignos. El cáncer puede darse en principio en cualquier órgano del cuerpo y a cualquier edad (American Cancer Society 2008; Jemal et al. 2005). Sin embargo, por dura que resulte esta realidad, la idea de que es una enfermedad incurable debe contemplarse como un mito obsoleto. La mayoría de los cánceres pueden tratarse, muchos de ellos pueden controlarse con éxito y algunos curarse por completo. Los índices de curación para algunas clases de cáncer llegan al 95% de los casos, por encima de los de algunas enfermedades infecciosas y desórdenes metabólicos.

Fundamentalmente el cáncer es un problema genético. Surge a partir de mutaciones y otras alteraciones patológicas en el genoma de una célula, induciendo a ésta y a sus descendientes a un comportamiento anómalo (Volgestein y Kinzler 2004). Estas alteraciones pueden heredarse en el momento de la concepción y afectar a todas las células de un organismo, pero por lo común se adquieren por accidente en un pequeño número de células en un tejido en particular. En la mayoría de los tipos de cáncer la transformación de una célula normal en cancerígena requiere múltiples mutaciones que, unidas, desactivan los mecanismos clave del autocontrol celular (fig. 1). Este cúmulo de mutaciones puede tardar décadas en producirse, una de las razones por la cual la incidencia del cáncer aumenta con la edad.

El cáncer es también un problema de biología celular. Las alteraciones genéticas que dan lugar al cáncer actúan alterando el ciclo de vida normal y el comportamiento social de las células (Gupta y Massagué 2006; Hanahan y Weinberg 2000). Los genes cuya función normal sería favorecer el movimiento y la división de células pueden convertirse en cancerígenos si sufren alteraciones que incrementen dichas actividades (fig. 2). Por otra parte, los genes cuya función normal es la de limitar la división celular, retener a las células en su sitio, favorecer la diferenciación celular o eliminar células muertas o defectuosas, conducen al cáncer si no son activados como requieren. La identificación de células cancerígenas y de

las funciones celulares que éstas controlan ocupa en la actualidad el primer plano de la investigación y del desarrollo de fármacos anticancerígenos.

La identificación de células cancerígenas y de sus funciones biológicas durante el último cuarto del siglo xx ha hecho posibles nuevas y mejores maneras de prevenir y tratar el cáncer. La mejora de los métodos de evaluación del riesgo de cáncer y de las campañas de prevención ha hecho disminuir la incidencia, así como la mortalidad, de ciertos tipos de cáncer. Procedimientos quirúrgicos menos invasivos, métodos de radioterapia más refinados y el perfeccionamiento de los medicamentos empleados en quimioterapia están contribuyendo al éxito creciente del tratamiento del cáncer convencional. Asimismo, una mejor comprensión de su biología y genética están posibilitando el desarrollo de nuevos y mejores medicamentos que tratan las células cancerígenas sin afectar a las sanas. Y aunque por ahora estos medicamentos aún llegan por goteo, vendrá el día en que este goteo se convertirá en inundación. La consecución de estos objetivos puede muy bien ser una de las grandes hazañas científicas del siglo xxi.

Creciente incidencia del cáncer

El cáncer no es una enfermedad nueva. Los egipcios ya lo trataban quirúrgicamente alrededor del 600 a.C. (Karpozilos y Pavlidis 2004). Hacia el 400 a.C. Hipócrates distinguió entre tumores benignos y malignos; a los segundos los llamó «carcinomas», a partir de la voz griega *carcinos*, que significa «cangrejo», en referencia a la forma que observó en los tumores malignos en estado avanzado, y el sufijo -oma, que significa «inflamación». Pero aunque no es una enfermedad nueva, la incidencia del cáncer va en aumento. Las estadísticas más recientes cifran la mortalidad anual por cáncer en casi ocho millones de personas, alrededor del 13% del total de muertes (Organización Mundial de la Salud 2008). La Organización Mundial de la Salud también predice que en 2020 esta cifra habrá ascendido a 11,5 millones.

Los tumores resultan de la acumulación de múltiples mutaciones en las células afectadas. Esta acumulación puede llevar años (Volgstein y Kinzler 2004). Por esta razón el cáncer es relativamente raro en niños y adolescentes y en cambio aumenta con la edad. En los paí-

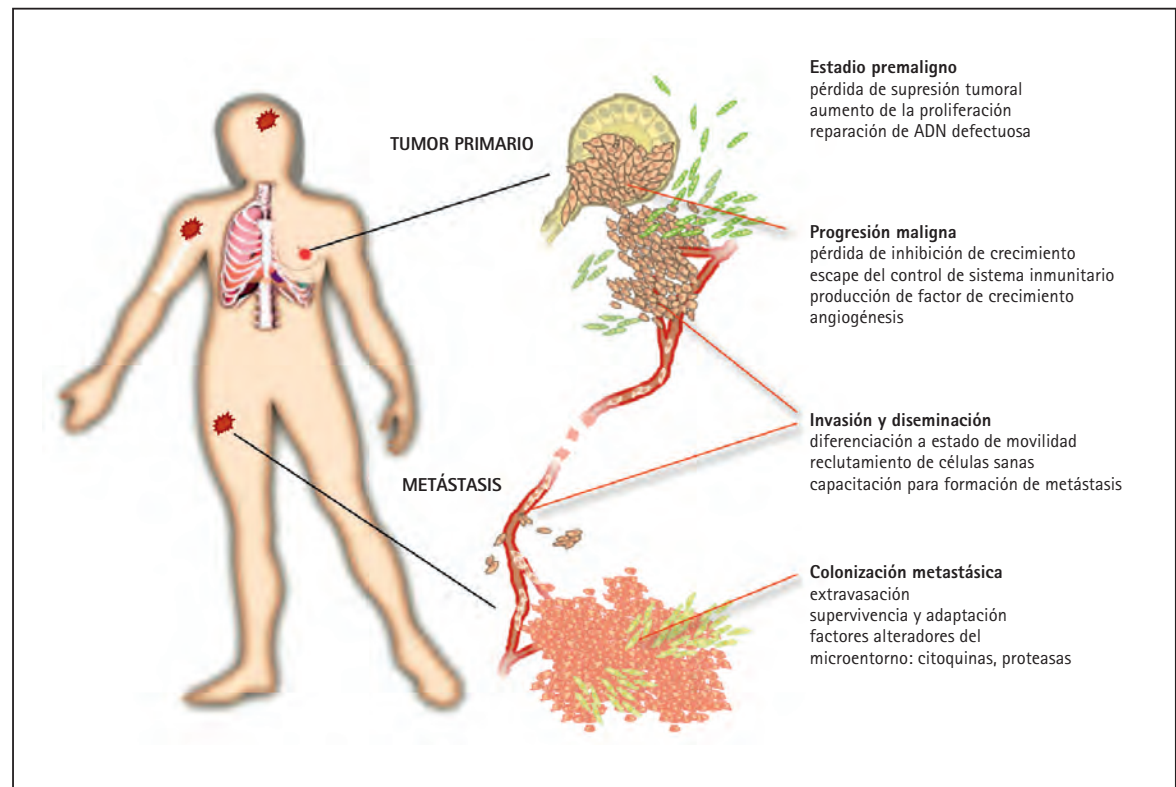


Figura 1. Fases de un tumor sólido. Los tumores sólidos tales como carcinomas de pulmón, colon, mama o próstata empiezan en las células epiteliales que se alinean en la superficie de los bronquios, de la mucosa intestinal o de los alveolos o secreción mucosa de las mamas y la próstata. Las mutaciones que aumentan la capacidad de progresión de estas células generan pequeñas masas pre-malignas de tejido. Estas lesiones pre-cancerosas pueden progresar y convertirse en tumores malignos si experimentan nuevas mutaciones que las liberen de los controles inhibidores de crecimiento, de la protección del sistema inmunitario y las doten de capacidad de invadir los tejidos adyacentes y de atraer capilares sanguíneos (angiogénesis). Una nueva conversión de los tumores malignos conduce a la formación de células cancerosas altamente invasivas y de gran movilidad y al reclutamiento de células sanas que ayudan a diseminar el tumor. Estos cambios preparan el camino para que las células cancerosas penetren el sistema linfático y la circulación sanguínea y lleguen a todas las partes del cuerpo. Algunas células cancerosas diseminadas pueden tener la capacidad de salirse de la circulación (extravasación) traspasando las paredes de los capilares sanguíneos. Una vez penetran órganos distantes como la médula ósea, los pulmones, el hígado o el cerebro, las células cancerosas pueden sobrevivir, adaptarse y finalmente conquistar estos nuevos entornos, dando lugar a la formación de metástasis letales.

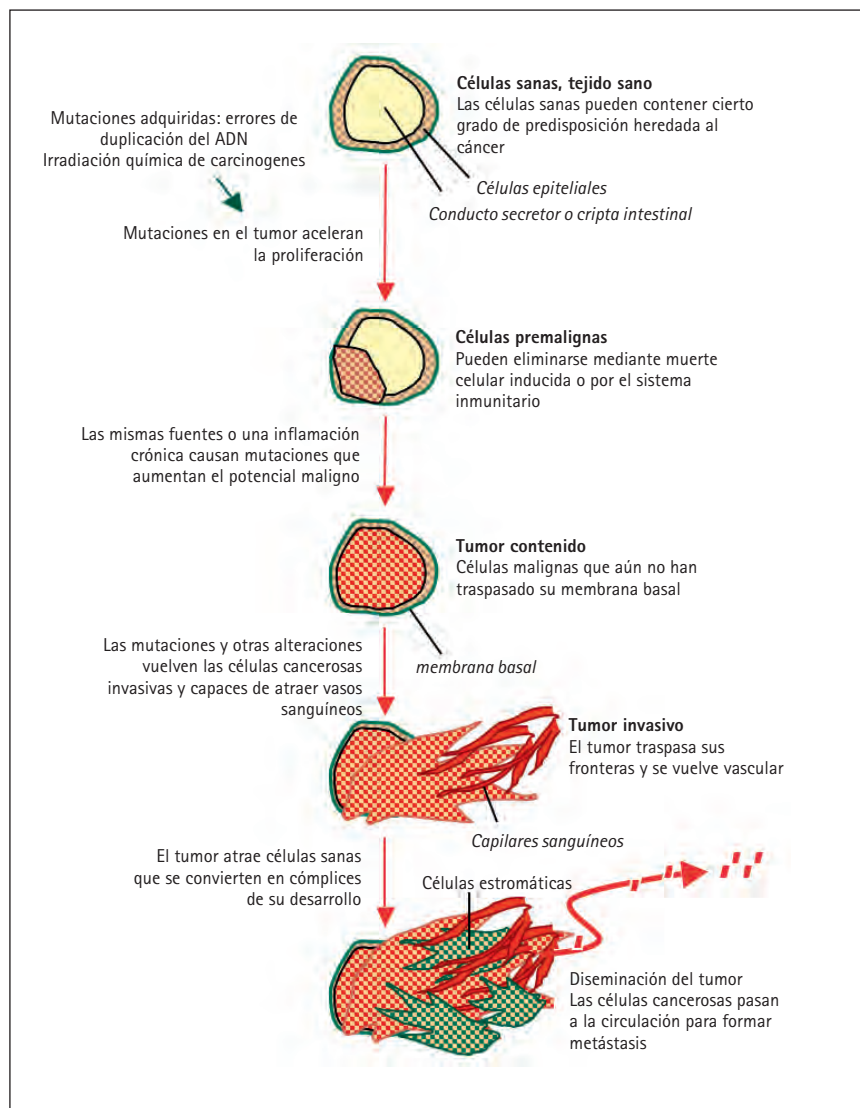


Figura 2. Origen de las mutaciones cancerosas. La ilustración representa la sección de un conducto secretor o cripta intestinal con una capa de células epiteliales rodeadas de una membrana basal recubriendo la cavidad. La herencia genética de cada individuo contiene un nivel determinado —alto o bajo— de predisposición a diferentes tipos de cáncer. Las variaciones genéticas de predisposición al cáncer que contienen un riesgo bajo de desarrollar un tipo determinado de cáncer son probablemente comunes en la población humana, y la mayoría están aún por descubrir. Las mutaciones heredadas que contienen un riesgo alto de desarrollar cáncer (por ejemplo, mutaciones BRCA1 y BRCA2 en cáncer de mama y ovario, mutaciones RB en retinoblastoma y mutaciones AP en carcinoma colorrectal) son poco frecuentes en la población. Estas predisposiciones intrínsecas al cáncer están presentes en todas las células del cuerpo. Sin embargo, el inicio de la formación de un tumor requiere, en todos los casos, que se produzcan más mutaciones. El origen de las mutaciones cancerosas puede ser interno, como errores no reparados en la replicación de ADN que las células sanas realizan por sí solas, o externo, como los carcinógenos químicos presentes en el tabaco o las radiaciones ultravioleta del sol. Estas mutaciones adquiridas aceleran la proliferación celular y conducen a la formación de lesiones pre-malignas, como pólipos intestinales o hiperplasias de tejido mamario. La mayoría de estas lesiones no progresan y son eliminadas por la autodestrucción celular o por el sistema inmunitario. Sin embargo, algunas lesiones pre-malignas pueden progresar hasta convertirse en un carcinoma in situ por la acumulación de mutaciones adicionales de origen externo o causadas por la inestabilidad genómica de las células pre-cancerosas. Esta progresión también puede producirse por síndromes de inflamación crónica desencadenados por un sistema inmunitario deficiente (por ejemplo, la colitis ulcerosa), un elemento irritativo externo (el tabaco en los pulmones) o un agente infeccioso (por ejemplo el virus de la hepatitis en el hígado, la bacteria *Helicobacter pylori* en el estómago). Un tumor se convierte en carcinoma invasivo cuando rompe la membrana basal que lo rodea y atrae a los vasos capilares para proveerse de oxígeno y nutrientes. Las alteraciones epigenéticas en las células cancerosas y el estrés en los tejidos adyacentes pueden causar la liberación de factores que reclutan células sanas, que terminan colaborando en la progresión del tumor. Llegada esta fase, las células cancerosas tienen acceso a la circulación y pueden diseminarse por el organismo. Algunas de estas células diseminadas pueden también reproducir el tumor en órganos distantes, causando la formación de metástasis.

ses desarrollados, el aumento de la esperanza de vida y de la población de mediana edad ocurrido en las últimas décadas ha contribuido a un crecimiento generalizado de la incidencia de cáncer. Dados los progresos conseguidos en el control de las enfermedades infecciosas que en la actualidad azotan a la población de los países en vías de desarrollo, cabe esperar que crezca también la incidencia de cáncer en ellos. Otros factores desencadenantes son la detección temprana de tumores en exámenes médicos rutinarios, factores relacionados con la dieta y los hábitos de vida y el impacto negativo del tabaco.

La incidencia del cáncer en general y la de determinados tipos de cáncer varía entre los países (Danaei et al. 2005; OMS 2008). Por ejemplo, los tipos de cáncer con mayores índices de mortalidad coinciden en Estados Unidos y España con una notable excepción: la mortalidad por cáncer de pulmón en las mujeres. Éste es el primer causante de muertes por cáncer entre la población masculina de ambos países. Sin embargo, al menos hasta hace poco tiempo, ocupaba el tercer puesto entre muertes por cáncer en mujeres en España (Cuadro 1). Esta diferencia se atribuye a que, comparadas con los hombres, las mujeres españolas empiezan a fumar más tarde en España que en Estados Unidos. Estudios epidemiológicos demuestran además una estrecha correlación entre tabaquismo y cáncer de pulmón, con una media de veinte años entre el inicio en el hábito y la aparición de la enfermedad.

Cáncer y cánceres

El término «cáncer» agrupa a cientos de enfermedades diferentes. Tumores primarios que se originan en diferentes órganos y tejidos —por ejemplo, cáncer de mama,

HOMBRES

Estados Unidos	España
pulmón (31%)	pulmón (28%)
próstata (10%)	colorectal (12%)
colorectal (8%)	próstata (8%)
páncreas (6%)	páncreas (6%)
leucemia (4%)	vejiga (6%)
hígado/vesícula biliar (4%)	estómago (6%)

MUJERES

Estados Unidos	España
pulmón (26%)	mama (19%)
mama (15%)	colorectal (15%)
colorectal (9%)	pulmón (8%)
páncreas (6%)	páncreas (6%)
ovario (6%)	ovario (6%)
leucemia (3%)	estómago (6%)

Cuadro 1. Incidencia de cáncer en adultos en Estados Unidos (American Cancer Society, 2008) y España (Centro Nacional de Epidemiología de España). Las cifras entre paréntesis representan el porcentaje total de muertes por ese tipo concreto de cáncer.

cáncer de pulmón o leucemia— son distintos entre sí en cuanto a apariencia, evolución, respuesta al tratamiento y mortalidad. Pero además, tumores originados en un mismo órgano pueden clasificarse en subclases que difieren mucho entre sí. Existen al menos cinco subclases de cáncer de mama, e incluso éstas podrían dividirse a su vez en diferentes variedades. Lo mismo puede decirse del cáncer en otros órganos. Estas diferencias requieren, pues, distintos tratamientos.

Los tumores también se clasifican de acuerdo al tipo de célula del que deriven. Los carcinomas son tumores malignos derivados de células epiteliales, tales como las que forman la capa superficial de la piel o epidermis, y en la mucosa digestiva o la estructura interna de órganos como la mama, la próstata, el hígado y el páncreas. Los sarcomas, por su parte, derivan de células de tejido conectivo como huesos, cartílagos y músculos. Los linfomas y las leucemias se originan en células precursoras; los melanomas en los melanocitos (células responsables de la pigmentación de la piel), y el glioblastoma, el neuroblastoma y el meduloblastoma, en células inmaduras del tejido neural. Los carcinomas son el tipo de cáncer más común en adultos, mientras que entre la población joven son más corrientes el neuroblastoma, el meduloblastoma y la leucemia.

Una tercera serie de parámetros en la clasificación de tumores se basa en su grado de invasión, lo que se conoce como *estadios* de la enfermedad, y su presentación histológica cuando se observan en el microscopio, llamada *grado*. Sin embargo, tumores de un mismo origen, tipo, grado y estadio pueden progresar y responder a la terapia de modos muy distintos dependiendo del paciente. Este hecho tiene un gran impacto en nuestra perspectiva del cáncer como enfermedad de la que aún sabemos muy poco. Por fortuna, esto está a punto de cambiar. La llegada de las tecnologías en genética molecular está haciendo posible una mejor clasificación de los tipos de cáncer basada en su origen específico, sus alteraciones moleculares, su riesgo de extenderse a otros órganos y su posible tratamiento.

Causas del cáncer

El cáncer se desarrolla como consecuencia de mutaciones y otras anomalías que alteran los genes que controlan el comportamiento celular (Hanahan y Weinberg 2000; Volgestein y Kinzler 2004). Estas mutaciones pueden producirse por la acción de factores externos —cancerígenos químicos, agentes infecciosos o radioactivos— o por errores internos en la replicación y corrección del ADN de pequeños grupos de células a lo largo del tiempo (fig. 2). Las mutaciones cancerígenas también pueden ser hereditarias, en cuyo caso se encuentran presentes en las células desde el momento de nacer. Las investigaciones actuales sobre genética se centran en los procesos que causan estas alteraciones, en los tipos de genes que se ven afectados y en sus consecuencias biológicas.

Ejemplos comunes de cancerígenos químicos incluyen el tabaco, que causa cáncer de pulmón y de vejiga, y la exposición a fibras de asbesto, que causa mesotelioma (Danaei et al. 2005). La radiación ultravioleta del sol puede producir melanoma y otros tipos de cáncer de piel. Se cree que los agentes cancerígenos presentes en el tabaco y las radiaciones favorecen la formación de tumores al actuar como mutágenos directos. El tabaco y los asbestos también pueden causar inflamaciones crónicas que a su vez pueden favorecer el desarrollo de tumores.

Las infecciones virales son la segunda causa externa más importante de cáncer después del tabaco (Zur Hausen 1999). Los virus asociados al cáncer en seres humanos incluyen el virus del papiloma en cáncer cervical, los virus de las hepatitis B y C en el cáncer de hígado, el VIH en el sarcoma de Kaposi y el virus de Epstein-Barr en linfomas de células B (Boshoff y Weiss 2002; Parato et al. 2005; Roden et al. 2006; Woodman et al. 2007; Young y Rickinson 2004). Las infecciones virales favorecen la formación de tumores al incorporar el genoma del virus en el ADN de la célula huésped, lo que puede incrementar la actividad de genes vecinos, los cuales a su vez estimulan la división incontrolada de células. Las infecciones virales también pueden favorecer la formación de tumores causando inflamación crónica y estimulando la producción de células en los tejidos huésped. La degeneración del tejido hepático, o cirrosis, causada por el alcoholismo, está relacionada con el desarrollo de cáncer de hígado. La combinación de cirrosis y hepatitis viral constituye el principal factor de riesgo de cáncer de hígado, que es de los más comunes y con mayor índice de mortalidad. El ejemplo más claro son los cánceres gástricos relacionados con la inflamación crónica de la mucosa estomacal por infección de la bacteria *Helicobacter pylori* (Cheung et al. 2007; Wang et al. 2007).

Ciertos tipos de cáncer tienen un fuerte componente hereditario (Volgestein y Kinzler 2004). Mutaciones heredadas en los genes BRCA1 y BRCA2 crean un componente de alto riesgo de desarrollar cáncer de mama y de ovarios (Walsh y King 2007; Wang 2007; Welsch y King 2001). Es interesante que las mutaciones en genes BRCA son poco frecuentes en cáncer esporádico (aquel en el que las mutaciones genéticas no son hereditarias, sino espontáneas). Por contra, el p53, que por lo general muta en los casos de cáncer esporádico, también es el gen afectado en el síndrome hereditario de Li-Fraumeni, que incluye predisposición a sarcomas, cáncer de mama y tumores cerebrales (Vousden y Lane 2007). El retinoblastoma en niños se debe a una mutación de carácter hereditario en el gen retinoblastoma (RB), que también muta en muchos tipos de cáncer esporádico (Classon y Harlow). Una mutación heredada del gen APC también puede dar lugar a la aparición de cientos de pólipos en el colon conducentes a un desarrollo temprano de carcinoma de colon (Fodde et al. 2001). Otra forma hereditaria de predisposición al cáncer es la causada por mutaciones

en uno de los varios genes (MLH1, MSH2, MSH6, PMS1, PMS2) dedicados a reparar errores de replicación en el ADN. Este trastorno hereditario (llamado HNPCC, cáncer de colon hereditario sin pólipos) incluye casos de cáncer de colon sin pólipos, cáncer uterino, gástrico y de ovarios dentro de una misma familia (De la Chapelle 2004). Mutaciones heredadas en el gen VHL1 producen una predisposición al cáncer de riñón (Kaelin 2005).

Las mutaciones hereditarias que tienen un fuerte efecto en el desarrollo del cáncer son raras entre los seres humanos y responsables sólo de una pequeña fracción de las estadísticas totales de cáncer. Por ejemplo, la mutación heredada del BRCA es responsable de menos de un 2% de los casos de cáncer de mama (Welsch y King 2001). En el extremo opuesto del espectro, ciertas variaciones genéticas pueden tener un impacto muy leve a escala individual en el riesgo de desarrollar cáncer, pero pueden ser prevalentes en los seres humanos. En determinados casos estos rasgos genéticos pueden combinarse para crear un riesgo significativo de cáncer. La visión predominante en la actualidad es que el cáncer surge a partir de complejas interacciones entre carcinógenos externos y el genoma individual. La identificación de estos determinantes genéticos en hoy día objeto de intensas investigaciones.

Células sanas y células cancerosas

La célula es la unidad básica de la vida. Aislada, sus actividades básicamente son resistir el entorno, incorporar nutrientes, replicar fielmente su genoma y dividirse. Sin embargo, las células que conforman los tejidos de un organismo complejo ya no son capaces de realizar estas tareas de forma autónoma. Células individuales evolucionaron y formaron colonias hace cientos de millones de años porque esta forma comunal de vida suponía una ventaja a la hora de enfrentarse a entornos hostiles. Pero también implicaba sacrificar cierto grado de libertad. Por ejemplo, ya no era posible para una célula dentro de una comunidad dividirse o desplazarse a voluntad. En nuestros tejidos altamente organizados, estas decisiones están sujetas a una complicada red de señales moleculares entre células. Esta forma de diálogo intercelular lleva desarrollándose y enriqueciéndose millones de años y una buena parte de nuestro genoma está enteramente dedicada a ella.

Las células se comunican unas con otras segregando moléculas, por lo general en forma de pequeñas proteínas llamadas hormonas, factores de crecimiento, citoquinas o quimiocinas. Estos factores contactan los receptores de proteínas en la superficie de las células de destino activando vías, que son secuencias de reacciones bioquímicas entre proteínas portadoras de señales dentro de la célula (Bierie y Moses 2006; Bild et al. 2006; Christofori 2006; Ciardello y Tortora 2008; Classon y Harlow 2002; Ferrara 2002; Fodde et al. 2001; Hanahan y Weinberg 2000; Karin 2006; Weinberg 2000; Malumbres y Barbacid 2007; Massagué 2004 y 2008; Olsson et al. 2006; Pouyssegur et al.

2006; Sweet-Cordero et al. 2005; Vousden y Lane 2007). El resultado final de este proceso son cambios positivos o negativos en la capacidad de la célula para desplazarse, metabolizar, crecer, dividirse, diferenciarse o morir. Otras proteínas del interior de la célula detectan la presencia de errores y alteraciones en el ADN y provocan, bien su reparación, bien su muerte. La pérdida de estas importantes funciones señalizadoras y de autocontrol deriva en la aparición de cáncer. Las células cancerosas desobedecen reglas esenciales de la vida en comunidad incrementando los estímulos proliferativos erróneos e ignorando las leyes de moderación. Su interacción con sus vecinas se vuelve abiertamente antisocial y escapan del control del sistema inmunitario. Con el tiempo rompen las barreras que encapsulan el tumor e inician un recorrido que diseminará las células cancerosas por el cuerpo, creando metástasis.

Las mutaciones causantes del cáncer afectan específicamente los genes encargados de ejercer estas funciones de control que tan importantes son. La acumulación progresiva de mutaciones convierte células normales en pre-malignas y, con el tiempo, en malignas (fig. 2). Estos cambios pueden observarse en el microscopio. Un proceso maligno puede empezar con la presencia de un número excesivo de células de apariencia normal, conocido como hiperplasia, y más específicamente con una acumulación desordenada de células de este tipo, llamada displasia. Cuando las células dejan de parecer normales la lesión se considera carcinoma in situ, en el cual las células anormales siguen estando dentro de los límites normales. Cuando las células del carcinoma invaden los tejidos adyacentes rompiendo su membrana o lámina basal, la lesión recibe el nombre de carcinoma invasivo. Cada una de estas etapas se acompaña de la progresiva acumulación de mutaciones que conducen al cáncer.

Las funciones específicas que deben ser perturbadas para que se generen células cancerígenas incluyen un aumento de autonomía en la emisión de señales inductoras de crecimiento; pérdida de sensibilidad a las señales inhibitorias de crecimiento; pérdida de la capacidad de muerte celular (llamada pérdida de apoptosis); aumento de la capacidad de replicar perpetuamente el ADN y aumento en la habilidad para escapar al control del sistema inmunitario (Hanahan y Weinberg 2000). Estos cambios son un requisito en todos los tipos de cáncer, incluidos los cánceres sanguíneos, la leucemia. Para formar un tumor, las células cancerosas procedentes de tejidos sólidos requieren además un aumento de la capacidad de resistir la hipoxia por medio de la inducción de nuevos capilares que alimenten el tumor (angiogénesis), así como el incremento de la capacidad de separarse e invadir los tejidos adyacentes (fig. 2). Para extender el tumor a distintos puntos dentro del organismo, las células cancerosas deben también adquirir la capacidad de pasar al sistema circulatorio, penetrar tejidos distantes y adaptarse al microentorno de dichos tejidos hasta terminar por apoderarse de ellos.

Genes del cáncer

Los genes del cáncer se dividen en dos grandes grupos. Aquellos cuyo exceso de actividad contribuye a la aparición de cáncer se denominan *oncogenes* (Hanahan y Weinberg 2000). Los genes codifican receptores de factores de crecimiento tales como el EGFR y el HER2, transductores de señales de crecimiento como RAS, RAF y P13K, factores de supervivencia celular como el BCL2 y otros. Las mutaciones que afectan a estos genes son activadoras o de «ganancia de función». Los genes cuya actividad normal previene la aparición del cáncer reciben el nombre de «supresores tumorales» y las mutaciones que los afectan en procesos cancerígenos son inactivadoras. Los supresores tumorales incluyen sensores del daño en el ADN como el p53, genes que reparan los daños en el ADN como los BRCA1 y BRCA2, inhibidores del ciclo de división celular como el RB, receptores y transductores de señales inhibitoras del crecimiento como el TGFBR y el SMAD4 y supresores de señales de crecimiento como el PTEN.

Las mutaciones que afectan a estos genes pueden ser puntuales, es decir, que afectan a un solo nucleótido del gen y a un solo aminoácido en el producto del gen. Las mutaciones puntuales pueden aumentar o reducir la actividad del producto del gen, y por tanto son una causa de activación oncogénica así como de la inactivación de genes supresores tumorales. Pequeñas pérdidas o inserciones también pueden causar activación oncogénica o inactivación de los supresores tumorales. Las mutaciones a gran escala incluyen pérdida o adquisición de una porción de cromosoma que resulta en la multiplicación de uno o más oncogenes, o una pérdida de genes supresores tumorales. Las translocaciones se producen cuando dos regiones cromosómicas diferenciadas se fusionan de forma irregular, a menudo en una localización determinada. Un ejemplo conocido de esto es el cromosoma Filadelfia o translocación de los cromosomas 9 y 22, que se da en la leucemia mieloide crónica y resulta en la producción de la proteína de fusión BCR-ABL (Melo y Barnes 2007). Ello causa activación oncogénica del gen ABL. Algunas mutaciones oncogénicas afectan no sólo a la región codificadora de proteínas de un oncogén sino a la región reguladora o «promotora» encargada de controlar el producto del gen. La inserción de un genoma viral cerca de la región promotora también puede llevar a la hiperactivación de un oncogén.

Además de las distintas clases de mutaciones que pueden alterar la estructura química de un gen normal convirtiéndolo en cancerígeno, los investigadores son cada vez más conscientes del impacto de las modificaciones epigenómicas. Éstas son modificaciones químicas del ADN y de las proteínas que lo rodean (Blasco 2007; Esteller 2007). Dichas modificaciones se conocen como cambios epigenéticos y tienen la capacidad de silenciar la expresión de un gen o de impedir que sea activado. La desregulación epigenética puede contribuir a la aparición de cáncer si no consigue silenciar la expresión de un gen

o hacer que sea competente para su activación. La pérdida de metilación puede desembocar en una expresión aberrante de oncogenes. La metilación o acetilación de proteínas histonas que envuelven el ADN cromosómico también pueden sufrir alteraciones que favorezcan el cáncer. El fármaco experimental anti cancerígeno Vorinostat actúa restaurando la acetilación de las histonas y está actualmente en fase de prueba.

Ecología del microentorno de los tumores

Cada tejido tiene una estructura, unos límites y una vascularización característicos, además de un entorno extracelular de hormonas, nutrientes y metabolitos. Las células cancerosas que alteran estas condiciones están expuestas a fuentes de estrés medioambiental incluyendo falta de oxígeno (hipoxia) y nutrientes, acidez, estrés oxidativo y respuestas inflamatorias. Las células que sobreviven a estos agentes de desgaste se convierten en población dominante en el tumor en desarrollo. Este fenómeno se conoce como «selección clonal» (Nowell 1976). Los clones de células resultantes no son meros supervivientes sino oportunistas altamente efectivos que se benefician del microentorno del tumor.

Los tumores son más que un conglomerado de células cancerígenas, también incluyen células sanas que son atraídas y finalmente engullidas por el creciente tumor y pueden convertirse en cómplices de su desarrollo (Joyce 2005; Mueller y Fusenig 2004). El conjunto de tipos de células no cancerígenas presentes en un tumor se conoce como *estroma* y su importancia en el cáncer es cada vez más reconocida. Células endoteliales presentes en el tumor forman nuevos capilares sanguíneos que atraen nutrientes y oxígeno a la masa tumoral. Macrófagos y otras células inmunes e inflamatorias se congregan en el tumor en un intento por responder al daño infringido a los tejidos. Los macrófagos asociados a tumores producen factores de crecimiento y enzimas ECM degradadas que estimulan el crecimiento y la invasión de células cancerosas (Joyce 2005; Lewis y Pollard 2006). El tumor también recluta células de defensa ante el estrés del sistema circulatorio. Varios tipos de células de la sangre son atraídas por señales que emanan del tumor y proliferan en respuesta a ellas. Los factores derivados del estroma pueden a su vez estimular a las células cancerosas a liberar señales que refuerzan su capacidad de producir metástasis. Por ejemplo, la citoquina derivada de estroma factor de crecimiento transformante tipo b (TGF- β) puede inducir a las células de cáncer de mama a liberar angiopoyetina-like 4, que refuerza la capacidad de estas células de invadir los pulmones antes de escapar del tumor primario (Padua et al. 2008). Así, el estroma de un tumor puede proporcionar ventajas metastásicas a las células cancerosas.

Metástasis: la extensión letal de los tumores

Los tumores agresivos pueden enviar millones de células cancerosas al sistema circulatorio antes de ser detec-

tados y extirpados quirúrgicamente. La metástasis es el proceso por el cual estas células cancerosas diseminadas invaden distintos órganos y terminan por causar disfunción en los mismos y, en última instancia, la muerte (fig. 1). Las metástasis pueden detectarse coincidiendo con el diagnóstico inicial o meses o años más tarde, cuando se produce una recidiva. Las células cancerosas diseminadas pueden permanecer en estado letárgico en distintos órganos durante largos periodos de tiempo, hasta que, por causas que se desconocen, se reactivan y empiezan a formar metástasis de crecimiento agresivo.

La administración de quimioterapia a pacientes de cáncer después de haberles extirpado un tumor primario tiene por objeto eliminar todas las células tumorales residuales y evitar la formación de metástasis. Sin embargo, el fracaso de las terapias existentes en la actualidad de curar metástasis es responsable del 90% de las muertes por cáncer. Si no fuera por las metástasis, el cáncer supondría únicamente una pequeña fracción del problema que es hoy en día. La comprensión de los múltiples procesos moleculares que participan en la formación de metástasis puede con el tiempo conducir a formas más efectivas de tratar esta enfermedad.

Los recientes avances tecnológicos a la hora de identificar y rastrear las metástasis han ayudado a perfilar los múltiples procesos que llevan a las células cancerosas de un tumor primario a alcanzar y colonizar órganos distantes (Fidler 2003; Gupta y Massagué 2006; Weinberg 2007) (fig. 2). Las células del carcinoma deben primero traspasar la membrana basal del tejido en que se encuentra el tumor. La membrana basal separa el epitelio de la célula en el que se ha originado el tumor del tejido subyacente. Las membranas basales también envuelven los vasos sanguíneos. Para traspasar la membrana y extenderse por el tejido adyacente las células cancerosas deben adquirir la capacidad de separarse de su lugar de origen, adoptar un comportamiento migratorio y liberar enzimas proteolíticas que degraden el armazón proteínico de la membrana basal y de la matriz extracelular.

Una vez las células cancerosas han formado una pequeña masa tumoral y creado condiciones de hipoxia, responden a ésta mediante la secreción de citoquinas, las cuales estimulan la formación de nuevos capilares que las proveen del oxígeno y los nutrientes necesarios para el crecimiento del tumor. Como resultado de estos factores de permeabilidad derivados del tumor, estos capilares nuevos son porosos y permiten que las células cancerosas escapen y entre en la circulación sanguínea. Los vasos linfáticos que filtran fluidos del tumor al tejido adyacente proporcionan una nueva ruta para la diseminación de células cancerosas. Los nódulos linfáticos a menudo atrapan células cancerosas en circulación y documentan su expansión, por eso el estado de los nódulos linfáticos constituye un importante indicador de pronóstico en el diagnóstico inicial. Sin embargo, la diseminación de

células cancerígenas a órganos distantes tales como los pulmones, el cerebro, los huesos y el hígado se produce principalmente a través de la circulación sanguínea. En ella las células cancerosas se asocian unas con otras y con células sanguíneas para formar embolias que pueden contribuir a resistir la tensión mecánica o a escapar de la vigilancia del sistema inmunitario.

Una vez las células cancerígenas se alojan en los capilares de órganos distantes deben atravesar las paredes capilares para acceder a la parénquima del órgano (fig. 3). La extravasación, como se conoce este proceso, depende de la capacidad de las células cancerosas de alterar los estrechos contactos existentes entre las células endoteliales de la pared capilar y de la membrana basal. El microentorno del órgano infiltrado es en gran medida hostil a las células cancerosas extravasadas, muchas de las cuales mueren. Aquellas que sobreviven forman micrometástasis que deben adaptarse al nuevo entorno y modificar las células residentes en él para reiniciar el proceso de crecimiento del tumor y formar colonias de metástasis agresivas. Este proceso puede llevar meses, años e incluso décadas. Tan sólo una pequeña fracción de las células cancerosas liberadas por un tumor cumple todos estos requisitos, pero las pocas que lo hacen bastan para que se formen metástasis letales.

Componentes de la metástasis

Heterogeneidad genética

La formación de metástasis tiene mucho de proceso evolutivo darwiniano: de una población de células cancerosas sólo las más fuertes sobreviven a las condiciones de su entorno. La evolución requiere de la presencia de heterogeneidad genética en una población en la que los individuos más fuertes pueden ser seleccionados para sobrevivir bajo determinadas presiones ambientales. En los tumores, dicha heterogeneidad viene garantizada por la inestabilidad genómica inherente a las células cancerosas, y aumenta la probabilidad de que algunas de las células de un tumor adquieran competencia metastásica. La integridad del ADN puede verse comprometida por aberraciones en la progresión del ciclo celular, crisis teloméricas, desactivación de los genes reparadores de ADN y alteración de los mecanismos de control epigenético. Por ejemplo, la mitad de los cánceres en seres humanos sufren pérdida del supresor tumoral p53, una proteína interna que responde a los daños en el ADN causando la eliminación de la célula dañada. La pérdida de p53 permite que las células cancerosas con alteraciones en el ADN sobrevivan y experimenten sucesivas mutaciones (Halazonetis et al. 2008). Mutaciones heredadas en determinados genes reparadores de ADN están asociadas a un riesgo mayor de desarrollar cáncer. Es el caso del síndrome de cáncer colorrectal sin poliposis hereditario (Rustgi 2007) y de cánceres de mama causados por mutaciones en el BRCA1 o el BRCA2 (Marin et al. 2008).

Células madre cancerosas

Las células cancerosas pueden diseminarse a partir de un tumor en estadios muy tempranos del mismo. Se han detectado células cancerosas en la médula ósea de pacientes con pequeños tumores de mama (Klein et al. 2002; Schmidt-Kittler et al. 2003). Ello no significa necesariamente que las primeras células migratorias sean las que progresan hasta convertirse en metástasis, pero sí indica que la diseminación no es propiedad exclusiva de tumores grandes y en estado avanzado.

Una vez las células diseminadas alcanzan órganos distantes pueden permanecer en estado letárgico e incluso morir. El estado letárgico puede durar años, incluso décadas hasta que las células cancerosas diseminadas inicien un crecimiento agresivo, como en el caso del cáncer de mama.

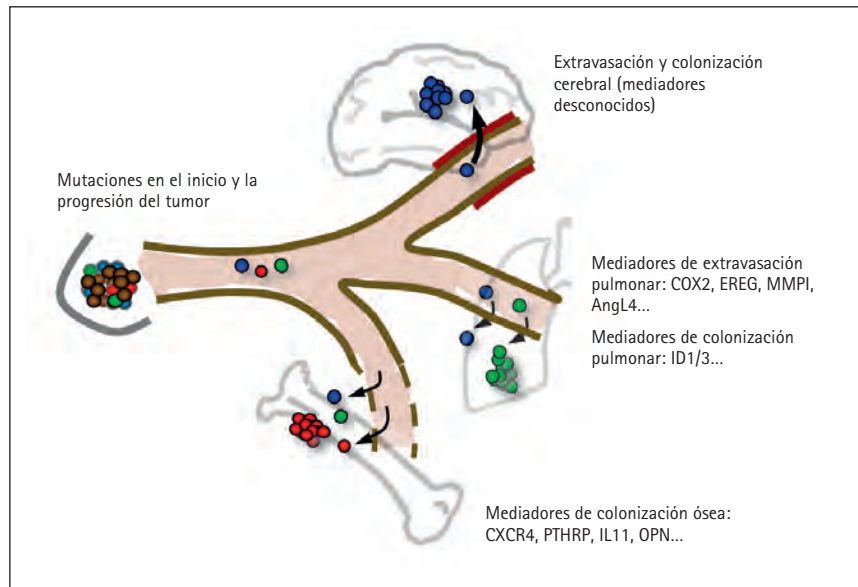


Figura 3. Mediadores en la formación de metástasis distantes en el cáncer de mama. He elegido para mi estudio las metástasis producidas por cáncer de mama debido a la alta incidencia de este tipo de tumores, la disponibilidad de material clínico y el número de órganos que pueden verse afectados. Los tumores mamarios pueden liberar células en la circulación sanguínea tan pronto como se vuelven localmente invasivos, experimentando mutaciones que favorecen el origen y la progresión del tumor (fig. 2). Las células diseminadas que sobreviven al estrés físico de la circulación requieren de nuevas funciones para penetrar tejidos distantes. El paso a dichos tejidos atravesando las paredes de los capilares sanguíneos o extravasación es relativamente permisivo en la médula ósea (y en el hígado, aunque no se ilustra aquí), porque los capilares de estos tejidos poseen ventanas naturales que permiten la entrada y salida constantes de células sanguíneas. Sin embargo, una vez penetran la médula ósea las células cancerosas deben ser capaces de sobrevivir e interactuar de forma productiva con este microentorno. El hecho de que las metástasis causadas por cáncer de mama pueden tardar años e incluso décadas en aparecer sugiere que las células cancerosas diseminadas llegaron originalmente a este órgano sin estar preparadas y tuvieron que desarrollar poco a poco las capacidades necesarias para expandirse en forma de agresivas colonias. Los genes de los que hacen uso las células de mama cancerosas para sobrevivir en la médula ósea incluyen el regulador de la quimiotaxis CXCR4, la proteína relacionada con la hormona paratiroidea estimuladora de osteoclastos PTHrP, el interleuquin-11 (IL11), el osteopontín (OPN) y otros genes. Al contrario de lo que ocurre con los capilares de la médula ósea, los capilares de otros órganos como los pulmones y en especial el cerebro tienen paredes resistentes que restringen el paso de las células en circulación. Para penetrar en estos órganos las células cancerosas deben por tanto transportar determinados genes activados. Entre ellos están el ligando epipegulina EGFR (EREG), la enzima ciclooxigenasa-2 inhibidora de prostaglandina (COX2), la enzima metaloproteínasa-1 degradadora de colágeno (MMP1) y el factor alterador del endotelio angipoyetina-like 4 (AngL4). Se sospecha que algunos de estos genes también contribuyen a la penetración de células de mama cancerosas en el cerebro. Los genes que hacen posible la colonización de los pulmones y el cerebro son en gran medida desconocidos y sujeto de activas investigaciones. El ID1 y el ID3 han sido identificados recientemente como mediadores en la reiniciación de tumores por la penetración de células de mama cancerosas en el cerebro. Así, la expresión del ID1/3 es una propiedad de las células propagadoras de tumores, también conocidas como células madre cancerosas.

Células cancerosas diseminadas encontradas en la médula ósea de mujeres o de ratones transgénicos con cánceres en estadios tempranos pueden activarse por trasplantes de médula ósea y causar tumores letales (Huseman et al. 2008).

La diseminación también puede darse en tumores metastásicos, que a su vez generan nuevas metástasis. Es posible que las células tumorales en circulación regresen a los mismos tumores de los que salieron. De acuerdo con esta hipótesis los tumores pueden alimentarse constantemente de su progenie más agresiva, proporcionando un mecanismo que aúna capacidad metastásica con crecimiento tumoral (Norton y Massagué 2006). Esto explicaría la persistente correlación entre metástasis y tamaño del tumor (Minn et al. 2007). El ritmo y los mecanismos de la diseminación de células cancerosas son objeto de gran atención en la investigación científica actual sobre cáncer.

Diferentes «semillas» para «suelos» distintos

Los huesos, los pulmones y el cerebro son las localizaciones más frecuentes de metástasis (fig. 3). Sin embargo, los diferentes tipos de cáncer son proclives a extenderse a diferentes órganos (Billingsley et al. 1999; Gavrilovi y Posner 2005; Hess et al. 2006; Leiter et al. 2004). La hipótesis sobre compatibilidad entre células cancerosas diseminadas (la «semilla») y determinados órganos distantes («suelos») ya la formuló en el siglo XIX Stephen Paget (Paget 1889). Así por ejemplo, el cáncer de mama puede extenderse a estos cuatro órganos, aunque los huesos y los pulmones son los más frecuentemente afectados. Las metástasis de cáncer de pulmón se localizan sobre todo en el cerebro, los huesos y el otro pulmón. Por contra, las metástasis de cáncer de próstata se dan sobre todo en los huesos y, en mucha menor medida, en los pulmones. Es más, aunque estos tres tumores se extienden a los huesos, en ellos forman lesiones muy distintas: las metástasis de hueso por cáncer de mama y de pulmón son osteolíticas, es decir, que se disuelven en la matriz ósea causando fracturas. En cambio la metástasis de cáncer de próstata es osteoblástica, es decir, que genera tejido óseo anormal que llena la cavidad medular. La preferencia de un tumor en un órgano por hacer metástasis dentro de ese mismo órgano también varía: los tumores en un pulmón hacen fácilmente metástasis en el otro, mientras que esto rara vez ocurre en el cáncer de mama.

Hacia una comprensión de las metástasis

Los progresos científicos que han acompañado la llegada del siglo XXI hacen posible una nueva visión de las metástasis basada en una mejor comprensión de sus fundamentos genéticos, moleculares y biológicos. Este conocimiento se acumula a gran velocidad a partir de la identificación de genes cuya actividad aberrante favorece la aparición de células metastásicas. Gracias a estos avances la metástasis está pasando de ser un oscuro objeto de estudio a un problema susceptible de analizarse racionalmente, diseccionarse y, en última instancia, resolverse.

Un modelo integrado de metástasis

Las primeras teorías sobre las metástasis proponían modelos antagónicos de predeterminación genética de una masa tumoral a metastatizar frente a la progresión del tumor resultante en la aparición de células anormales susceptibles de formar metástasis (Volgestein et al. 1998). Con la secuenciación del genoma humano se han desarrollado potentes tecnologías de *microarray* (microconfiguración) que permiten a los investigadores determinar el estado de actividad de cada gen de una pequeña muestra de tejido. Empleando dichas técnicas ha sido posible identificar patrones de actividad génica o «perfiles de expresión génica», capaces de indicar las probabilidades de que un tumor particular cause metástasis. Si una muestra extraída de un tumor primario revela la presencia de un perfil de expresión génica pro metastático, ello indicaría que una proporción sustancial de las células de dicho tumor están expresando genes de este tipo y por tanto reúnen las condiciones —es decir, son competentes— para que se formen metástasis. Ello apoyaría la teoría sobre la predeterminación de las metástasis. Sin embargo, esta supuesta competencia de los genes puede ser incompleta. Han de producirse alteraciones adicionales antes de que las células cancerosas estén completamente equipadas para invadir y colonizar un tejido distante. La adquisición de todas las condiciones necesarias para que se formen metástasis puede darse de forma masiva en la población de un tumor, como es el caso de tumores que metastatizan rápidamente en múltiples órganos, o bien puede producirse lentamente en una minoría de células especialmente predispuestas, dando lugar a metástasis en uno u otro órgano, años o incluso décadas después de haber salido del tumor primario. Este último caso confirmaría la teoría de la progresión del tumor como un requisito necesario para la formación de metástasis.

Progresos recientes en el estudio de las metástasis han proporcionado pruebas experimentales y clínicas para ambos modelos, el de predeterminación y el de progresión, con el resultado de la formulación de un tercer modelo que integra los dos. Las células cancerosas de un tumor con prognosis desfavorable pueden contener genes activados que las doten de algunas, aunque no todas, las funciones requeridas para la formación de metástasis distantes. A estos genes los llamamos genes *de progresión metastática*, porque permiten directamente a la población de células cancerosas adquirir la competencia necesaria para que se dé el comportamiento metastático. Los genes de progresión metastática son necesarios pero no bastan para crear metástasis, porque la mayoría de las células cancerosas que expresan dichos genes son todavía incapaces de formar tumores metastáticos. Esto implica la existencia de un conjunto complementario de genes metastáticos que proporcionan funciones adicionales de supervivencia y adaptación en un órgano determinado. A estos genes los llamamos *de virulencia metastática*.

Genes de progresión metastática

Recientes trabajos realizados en nuestro laboratorio han identificado un conjunto de dieciocho genes que utilizan las células de mama cancerosas para proliferar tanto en el tumor primario como en los pulmones (fig. 3). Este conjunto, al que llamamos «perfil de expresión génica de metástasis de pulmón» o LMS incluye EREG, COX-2 y MMP1, los cuales cooperan en al remodelación de nuevos capilares sanguíneos en tumores mamarios y de los ya existentes en los pulmones cuando las células cancerosas favorecen la unión de capilares porosos que facilitan la salida de células cancerosas; en el pulmón estos mismos productos facilitan el paso de células cancerosas en circulación en la parénquima (Gupta et al. 2007). Otro ejemplo es el gen que codifica el ID1, que inhibe la diferenciación celular y estabiliza la capacidad de las células cancerosas de propagar el tumor. En modelos experimentales el ID1 es importante para el crecimiento de tumores de mama y para la reiniciación del crecimiento del tumor una vez las células cancerosas han alcanzado los pulmones. Así, los genes de progresión metastática pueden cumplir los requisitos en cuanto a tejido del microentorno en un órgano particular para influir en la progresión del tumor primario. Pacientes de cáncer de mama con tumores primarios positivos en LMS tienen un riesgo mayor de desarrollar metástasis de pulmón, pero no de hueso o de otros órganos.

No todos los genes metastáticos que se expresan en tumores primarios proporcionan una ventaja selectiva en dichos tumores. Por ejemplo, la producción del factor transformador del crecimiento beta (TGFbeta) en el estroma de tumores primarios de mama estimula la expresión de más de cien genes en las células cancerosas mamarias del mismo tumor. Entre ellos está el gen que contiene el factor secretado ANGPTL4. A diferencia de EGFR, el COX2, el MMP1 o el ID1, la producción de ANGPTL4 no parece proporcionar ventaja ninguna a las células cancerosas en los tumores primarios, sino que únicamente refleja la presencia de TGFbeta en el tumor. Sin embargo, cuando las células cancerosas estimuladas llegan a los capilares del pulmón, el ANGPTL4 que secretan causa la rotura de la pared capilar y permite que el cáncer penetre en tejido (Padua et al. 2008).

Contribuciones específicas de los genes asociados a la virulencia metastática

Cuando las células cancerosas alcanzan órganos distantes por lo general se enfrentan a un microentorno no permisivo. Para formar una colonia de metástasis el cáncer debe ser capaz de resistir y explotar su microentorno. Claro ejemplo de ello son las metástasis osteolíticas de hueso causadas por el cáncer de mama. Las células cancerosas en circulación que penetran la médula ósea deben encontrar la forma de sobrevivir en el entorno hormonal y estrómic de este tejido, así como la manera de activar la movilización y acción de los osteoclastos que hacen posible la destrucción de los huesos. Las células de mama cancerosas

que forman metástasis de hueso muestran niveles elevados de CXCR4. Esta proteína de membrana actúa como receptora para el factor de supervivencia celular CXCL12, que se produce en grandes cantidades en el estroma de la médula ósea (Wang et al. 2006). Por tanto las células cancerosas que expresan niveles altos de CXCR4 obtienen una ventaja específica de la presencia de CXCL12 en la médula ósea.

En modelos experimentales que emplean ratones, las células de mama cancerosas que colonizan preferentemente los huesos muestran no sólo una elevada expresión del receptor de supervivencia celular CXCR4 sino también una alta producción de los factores PTHrP (péptido relacionado con la paratiroidea), TNF-alfa, IL-1, IL-6 e IL-11 (Kang et al. 2003). Cuando son secretados por células óseas metastásicas, estos factores estimulan a los osteoblastos para que liberen RANKL, el cual activa la diferenciación osteoblástica. Los osteoblastos disuelven el hueso y liberan factores de crecimiento como el IGF-1 o factor de crecimiento insulínico, que favorece la supervivencia de las células cancerosas, y TGFbeta, que estimula las células cancerosas para que continúen segregando PTHrP. El resultado final de este proceso es un ciclo vicioso de interacciones entre osteoclastos y células cancerosas que aceleran la acción destructiva de las metástasis óseas.

La búsqueda actual de genes y funciones que intervienen en la formación de metástasis en forma de tumores de otra clase o en otros órganos comienza a dar resultados. Las células de cáncer de próstata segregan factores como el Wnt y proteínas óseas morfogenéticas (BMP) que estimulan la acumulación de osteoblastos. Como resultado de ellos el cáncer de próstata da lugar a metástasis osteoblásticas (producción anormal de hueso), en contraste con las metástasis producidas por el cáncer de mama, que causan la destrucción de los huesos. Comparadas con las metástasis en hueso y pulmón es muy poco lo que sabe acerca de los genes que usan las células cancerosas para colonizar el hígado o el cerebro. Sin embargo este tema es objeto de intensas investigaciones que a buen seguro terminarán por dar fruto en un futuro no lejano.

Fronteras en la prevención, diagnóstico y tratamiento del cáncer

Las campañas de prevención del cáncer enfocadas a reducir los comportamientos de riesgo (consumo excesivo de tabaco y alcohol, exceso de exposición al sol y otros), así como los exámenes médicos rutinarios en los que se detectan tumores en estadios tempranos son cruciales para reducir la incidencia y la mortalidad de esta enfermedad. El diagnóstico precoz hace posible intervenir el tumor antes de que éste se disemine, curando al paciente o al menos alargando su esperanza de vida. Se ha avanzado mucho en la detección de cáncer de mama por medio de mamografías, de cáncer colorrectal con colonoscopias, de cáncer de útero mediante citología y de cáncer de próstata con exámenes rectales y análisis de sangre que miden los niveles

del PSA, o antígeno prostático específico. La vacunación preventiva contra el virus del papiloma humano de transmisión sexual tiene como fin reducir la incidencia del cáncer cervical. Se realizan exámenes genéticos para detectar ciertas mutaciones relacionadas con el cáncer en el BRCA1 y el BRCA2 (factores que predisponen a los cánceres de mama y ovarios) y para genes reparadores de ADN (que indican predisposición al cáncer de colon y otros) en individuos considerados de alto riesgo, con historial familiar de dichas enfermedades. Los portadores de estas mutaciones pueden así ser vigilados de cerca e incluso optar por cirugía profiláctica (mastectomías, extirpación de ovarios y de colon) para así reducir los riesgos de desarrollar el tumor.

Avances recientes están mejorando los enfoques clásicos en el tratamiento del cáncer (cirugía, quimioterapia, radioterapia) y los nuevos están basados en terapia dirigida e inmunoterapia. Los métodos quirúrgicos han ganado en precisión y son menos invasivos. La extirpación quirúrgica de un tumor que no se ha extendido puede curar el cáncer de forma efectiva. Sin embargo, la propensión de células cancerosas a invadir los tejidos adyacentes y extenderse a localizaciones distantes limita considerablemente la eficacia de la cirugía. Incluso tumores pequeños y localizados tienen potencial metastásico. Por tanto la cirugía se complementa a menudo con otras terapias. La radioterapia se basa en el uso de radiaciones ionizantes (rayos X) para encoger los tumores antes de la cirugía y eliminar células diseminadas localmente. La radiación puede dañar los tejidos sanos, por lo que sólo se aplica en áreas limitadas del cuerpo. La radioterapia destruye células causando extensos daños a su ADN. La mayoría de las células sanas puede recuperarse después de la radioterapia, lo que ofrece una ventana de posibilidad de que el tratamiento funcione.

La quimioterapia consiste en tratar el cáncer con fármacos que son más tóxicos para las células cancerosas que para las sanas. Los fármacos empleados tradicionalmente envenenan las células que se están dividiendo, interrumpiendo así la duplicación del ADN o la separación de cromosomas recién formados. Como en el caso de la radioterapia, las células normales tienen mayor capacidad de recuperarse del daño que las cancerosas. Por esta razón la quimioterapia a menudo se administra en las dosis máximas tolerables, con los consiguientes efectos secundarios relativos a la regeneración celular, tales como la pérdida de mucosa oral y gastrointestinal, cabello, piel y uñas.

Una meta importante de la investigación actual es desarrollar fármacos dirigidos a las células cancerosas pero basados en la dependencia de dichas células de las mutaciones oncogénicas que contienen (Sawyers 2004) (fig. 1). Estas *terapias dirigidas* no son diferentes de muchos fármacos empleados para otro tipo de enfermedades. Los fármacos dirigidos contra el cáncer buscan alcanzar mayor efectividad terapéutica con menos efectos secundarios. Su utilización permite que los niveles de radioterapia aplicados sean menores y menores también por tanto sus

efectos secundarios. Las terapias dirigidas incluyen compuestos químicos que por lo general actúan inhibiendo la actividad enzimática de los productos de los oncogenes, y anticuerpos monoclonales que actúan bloqueando el receptor oncogénico de la superficie de una célula o matando la célula cancerosa, destruyendo sus anticuerpos.

Las terapias dirigidas surgieron en la década de 1990 como consecuencia directa de la identificación de genes cancerosos críticos. La capacidad para analizar tumores molecularmente está revolucionando la clasificación, la prognosis y el tratamiento del cáncer. Las investigaciones en este campo continúan siendo intensas y prometedoras. Entre los anticuerpos monoclonales, el anticuerpo anti-HER2 trastuzumab (Herceptin™) resulta efectivo contra un subtipo de carcinomas de mama que contienen un oncogén HER2 activado (Hudis 2007; Shawver et al. 2002). El anticuerpo anti-CD20 rituximab (Rituxan™) se usa para tratar linfomas de células B que presentan el antígeno CD20 (Cheson y Leonard 2008), y el anticuerpo anti-EGFR cetuximab (Erbix™) se usa en cáncer de colon avanzado (Mendelsohn y Baselga 2006) (fig. 4).

Entre los compuestos químicos dirigidos, el imatinib (Gleevec™), un inhibidor de pequeñas moléculas del oncogén BCR-ABL quinasa se emplea con éxito contra leucemias causadas por este oncogén (Schiffer 2007). El inhibidor EGFR erlotinib (Tarceva™) se usa en carcinomas de pulmón causados por un EGFR mutante (Ciardello y Tortora 2008). Además, aunque diferentes subtipos de cáncer en un órgano determinado pueden experimentar clases de mutaciones diferentes, ciertos subtipos de cáncer en

órganos distintos pueden, sorprendentemente, compartir mutaciones. Como resultado de ello un mismo fármaco puede ser efectivo en tumores de tipo molecular localizados en órganos distintos. Una clase relacionada de compuestos anticáncer son los inhibidores de la angiogénesis, que previenen la formación de capilares sanguíneos que alimentan los tumores. Algunos, tales como el anticuerpo monoclonal bevacizumab (Avastin™) son ya de uso clínico (Ferrara 2002) (fig. 4), aunque con escaso éxito, debido a que las células cancerosas tienen múltiples maneras de estimular la angiogénesis (Meyerhardt y Mayer 2005). Las investigaciones actuales en este campo por tanto se centran en identificar combinaciones de inhibidores de la angiogénesis que sean efectivas (Bergers y Hanahan 2008).

Siguen siendo muchas las barreras para mejorar el tratamiento del cáncer, lo que conforma la necesidad de emprender nuevas direcciones. ¿Qué cambios podemos prever en el campo de la oncología? En un futuro no muy lejano, el perfil tumoral de un paciente podría incluir no sólo gradación histopatológica e información sobre el estado de las mutaciones oncogénicas comunes, sino también un completo retrato molecular del tumor (Massagué 2007; Nevins y Pottu 2007). Avances recientes en el diseño de perfiles moleculares de tumores han llevado al descubrimiento de perfiles de expresión de genes que permiten clasificar mejor los tumores en subtipos diferenciados, predecir los riesgos y la localización de metástasis con mayor precisión e identificar mejor los objetivos terapéuticos (Bild et al. 2006; Fan et al. 2006; Minn et al. 2005; Padua et al. 2008; Van't Veer et al. 2002, Van de Vijver et al. 2002). Un perfil de «prognosis desfavorable» de setenta genes (MammaPrint) y un conjunto de veintinueve genes no coincidentes de «recurrencia» (Oncotype Dx) han sido transformados en fármacos comerciales que ayudan a los médicos a tomar decisiones, evitando mastectomías en pacientes de cáncer de mama como tumores que presentan una prognosis favorable después de la quimioterapia. Los genes de dichos perfiles pueden ser entonces examinados directamente para medir su capacidad de formar metástasis y tratados con terapia dirigida para reducir la actividad metastásica de las células cancerosas (Gupta et al. 2007). Los tratamientos farmacológicos para pacientes de cáncer pueden incluir combinaciones individualizadas dirigidas a subtipos específicos de tumor y a sus metástasis. La mejora de los biomarcadores de respuesta a los fármacos en pacientes ayudará a evaluar con mayor precisión la reacción de éstos a terapias dirigidas (Sawyers 2008).

Los nuevos conocimientos acerca de las bases moleculares, genéticas y celulares del desarrollo y la progresión del cáncer traen consigo nuevas oportunidades de mejorar y expandir nuestra capacidad de prevenir, detectar y tratar esta enfermedad. Trabajando en estrecha colaboración, científicos y médicos pueden generar y aplicar los conocimientos necesarios para hacer del cáncer una enfermedad crónica más en las próximas décadas.

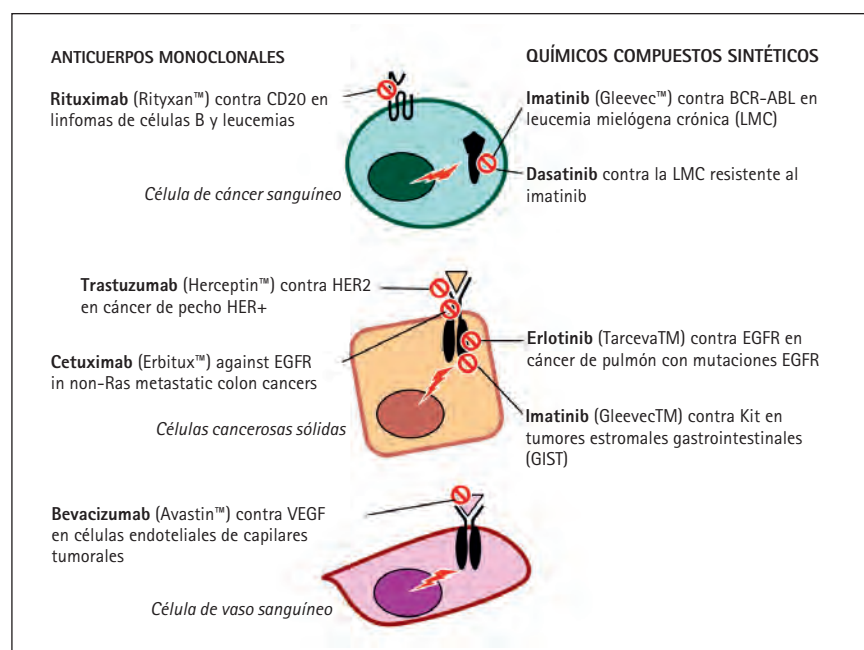


Figura 4. Terapia farmacológica dirigida contra el cáncer. La identificación de los genes y de ciertas proteínas que causan ciertos tipos de cáncer ha permitido desarrollar fármacos específicamente dirigidos a estas proteínas que bloquean su actividad, matando por tanto las células cancerosas que dependen de ellas para sobrevivir. Estos fármacos dirigidos incluyen anticuerpos monoclonales así como compuestos químicos sintéticos.

Bibliografía

- American Cancer Society. Cancer facts and figures 2008. <http://www.cancer.org/>, 2008.
- Bergers, G., y D. Hanahan. «Modes of resistance to anti-angiogenic therapy», *Nature Reviews Cancer* 8 (2008): 592-603.
- Bierie, B., y H. L. Mose. «Tumor microenvironment: TGFbeta: the molecular Jekyll and Hyde of cancer», *Nature Reviews Cancer* 6 (2008): 506-520. (*)
- Bild, A. H., G. Yao, J. T. Chang, Q. Wang, A. Potti, D. Chasse, M. B. Joshi, D. Harpole, J. M. Lancaster, A. Berchuck et al. «Oncogenic pathway signatures in human cancers as a guide to targeted therapies». *Nature* 439 (2006): 353-357.
- Billingsley, K.G., M. E. Burt, E. Jara, R. J. Ginsberg, J. M. Woodruff, D. H. Leung y M. F. Brennan. «Pulmonary metastases from soft tissue sarcoma: analysis of patterns of diseases and postmetastasis survival». *Annals of Surgery* 229, (1999): 602-610; *Discussion* (1999): 610-602.
- Blasco, M. A. «The epigenetic regulation of mammalian telomeres». *Nature Reviews Genetics* 8 (2007): 299-309.
- Boshoff, C. y R. Weiss. «AIDS-related malignancies». *Nature Reviews Cancer* 2 (2002): 373-382.
- Cano, A., M. A. Pérez-Moreno, I. Rodrigo, A. Locascio, M. J. Blanco, M. G. del Barrio, F. Portillo y M. A. Nieto. «The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression». *Nature Cell Biology* 2 (2000): 76-83.
- Cheson, B. D. y J. P. Leonard. «Monoclonal antibody therapy for B-cell non-Hodgkin's lymphoma». *The New England Journal of Medicine* 359 (2008): 613-626.
- Cheung, T. K., H. H. Xia y B. C. Wong. «Helicobacter pylori eradication for gastric cancer prevention». *Journal of Gastroenterology* 42, supl 17 (2007): 10-15.
- Christofori, G. «New signals from the invasive front». *Nature* 441 (2006): 444-450.
- Ciardiello, F. y G. Tortora, G. «EGFR antagonists in cancer treatment». *The New England Journal of Medicine* 358 (2008): 1.160-1.174.
- Clark, E. A., T. R. Golub, E. S. Lander y R. O. Hynes. «Genomic analysis of metastasis reveals an essential role for RhoC». *Nature* 406 (2002): 532-535.
- Clarke, M. F. y M. Fuller. «Stem cells and cancer: two faces of eve». *Cell* 124, (2006): 1.111-1.115.
- Classon, M. y E. Harlow. «The retinoblastoma tumour suppressor in development and cancer». *Nature Reviews Cancer* 2 (2002): 910-917.
- Danaei, G., S. Vander Hoorn, A. D. López, C. J. Murray y M. Ezzati. «Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors». *Lancet* 366 (2005): 1.784-1.793.
- De la Chapelle, A. «Genetic predisposition to colorectal cancer». *Nature Reviews Cancer* 4 (2004): 769-780.
- Esteller, M. «Cancer epigenomics: DNA methylomes and histone-modification maps». *Nature Reviews Genetics* 8 (2007): 286-298.
- Fan, C., D. S. Oh, L. Wessels, B. Weigelt, D. S. Nuyten, A. B. Nobel, L. J. Van't Veer y C. M. Perou. «Concordance among gene-expression-based predictors for breast cancer». *The New England Journal of Medicine* 355 (2006): 560-569.
- Ferrara, N. «VEGF and the quest for tumour angiogenesis factors». *Nature Reviews Cancer* 2 (2002): 795-803.
- Fidler, I. J. «The pathogenesis of cancer metastasis: the «seed and soil» hypothesis revisited». *Nature Reviews Cancer* 3 (2003): 453-458.
- Fodde, R., R. Smits y H. Clevers. «APC, signal transduction and genetic instability in colorectal cancer». *Nature Reviews Cancer* 1 (2001): 55-67.
- Gavrilovic, I. T. y J. B. Posner. «Brain metastases: epidemiology and pathophysiology». *Neuro-oncology* 75 (2005): 5-14.
- Gupta, G. P. y J. Massagué. «Cancer metastasis: building a framework». *Cell* 127 (2006): 679-695.
- , D. X. Nguyen, A. C. Chiang, P. D. Bos, J. Y. Kim, C. Nadal, R. R. Gomis, K. Manova-Todorova y J. Massagué. «Mediators of vascular remodelling co-opted for sequential steps in lung metastasis». *Nature* 446 (2007): 765-770.
- Halazonetis, T. D., V. G. Gorgoulis y J. Bartek. «An oncogene-induced DNA damage model for cancer development». *Science* 319 (2008): 1.352-1.355.
- Hanahan, D. y R. A. Weinberg. «The hallmarks of cancer». *Cell* 100 (2000): 57-70.
- Hess, K. R., G. R. Varadhachary, S. H. Taylor, W. Wei, M. N. Raber, R. Lenzi y J. L. Abbruzzese. «Metastatic patterns in adenocarcinoma». *Cancer* 106 (2006): 1.624-1.633.
- Hudis, C. A. «Trastuzumab-mechanism of action and use in clinical practice». *The New England Journal of Medicine* 357 (2007): 39-51.
- Husemann, Y., J. B. Geigl, F. Schubert, P. Musiani, M. Meyer, E. Burghart, G. Forni, R. Eils, T. Fehm, G. Riethmuller et al. «Systemic spread is an early step in breast cancer». *Cancer Cell* 13 (2008): 58-68.
- Jemal, A., T. Murray, E. Ward, A. Samuels, R. C. Tiwari, A. Ghafoor, E. J. Feuer y M. J. Thun. «Cancer statistics, 2005». *CA: a cancer journal for clinicians* 55 (2005): 10-30.
- Joyce, J. A. «Therapeutic targeting of the tumor microenvironment». *Cancer Cell* 7 (2005): 513-520.
- Kaelin, W. G. «The von Hippel-Lindau tumor suppressor protein: roles in cancer and oxygen sensing». *Cold Spring Harbor Symposia in Quantitative Biology* 70 (2005): 159-166.
- Kang, Y., P. M. Siegel, W. Shu, M. Drobnjak, S. M. Kakonen, C. Cordon-Cardo, T. A. Guise y J. Massagué. «A multigenic program mediating breast cancer metastasis to bone». *Cancer Cell* 3 (2003): 537-549.
- Karin, M. «Nuclear factor-kappaB in cancer development and progression». *Nature* 441 (2006): 431-436.
- Karpozilos, A. y N. Pavlidis. «The treatment of cancer in Greek antiquity». *European Journal of Cancer* 40 (2004): 2.033-2.040.
- Kim, M., J. D. Gans, C. Nogueira, A. Wang, J. H. Paik, B. Feng, C. Brennan, W. C. Hahn, C. Cordon-Cardo, S. N. Wagner et al. «Comparative oncogenomics identifies NEDD9 as a melanoma metastasis gene». *Cell* 125 (2006): 1.269-1.281.
- Klein, C. A., T. J. Blankenstein, O. Schmidt-Kittler, M. Petronio, B. Polzer, N. H. Stoecklein y G. Riethmuller. «Genetic heterogeneity of single disseminated tumour cells in minimal residual cancer». *Lancet* 360 (2002): 683-689.
- Leiter, U., F. Meier, B. Schitteck y C. Garbe. «The natural course of cutaneous melanoma». *Journal of Surgical Oncology* 86 (2004): 172-178.
- Lewis, C. E. y J. W. Pollard. «Distinct role of macrophages in different tumor microenvironments». *Cancer Research* 66 (2006): 605-612.
- Malumbres, M. y M. Barbacid. «Cell cycle kinases in cancer». *Current Opinion in Genetics and Development* 17 (2007): 60-65.
- Martin, R. W., P. P. Connell y D. K. Bishop. «The Yin and Yang of treating BRCA-deficient tumors». *Cell* 132 (2008): 919-920.
- Massagué, J. «G1 cell-cycle control and cancer». *Nature* 432 (2004): 298-306.

- , «Sorting out breast-cancer gene signatures». *The New England Journal of Medicine* 356 (2007): 294-297.
- , «TGFbeta in Cancer». *Cell* 134 (2008): 215-230.
- Melo, J. V. y D. J. Barnes. «Chronic myeloid leukaemia as a model of disease evolution in human cancer». *Nature Reviews Cancer* 7 (2007): 441-453.
- Mendelsohn, J. y J. Baselga. «Epidermal growth factor receptor targeting in cancer». *Seminars in Oncology* 33 (2006): 369-385.
- Meyerhardt, J. A. y R. J. Mayer. «Systemic therapy for colorectal cancer». *The New England Journal of Medicine* 352 (2005): 476-487.
- Minn, A.J., Gupta, G.P., Padua, D., Bos, P., Nguyen, D.X., Nuyten, D., Kreike, B., Zhang, Y., Wang, Y., Ishwaran, H., et al. (2007). Lung metastasis genes couple breast tumor size and metastatic spread. Proceedings of the National Academy of Sciences USA 104, 6740-6745.
- Minn, A. J., G. P. Gupta, P. M. Siegel, P. D. Bos, W. Shu, D. D. Giri, A. Viale, A. B. Olshen, W. L. Gerald y J. Massagué. «Genes that mediate breast cancer metastasis to lung». *Nature* 436 (2005): 518-524.
- Mueller, M. M. y N. E. Fusenig. «Friends or foes - bipolar effects of the tumour stroma in cancer». *Nature Reviews Cancer* 4 (2004): 839-849.
- Nevins, J. R. y A. Potti. «Mining gene expression profiles: expression signatures as cancer phenotypes». *Nature Reviews Genetics* 8 (2007): 601-609.
- Norton, L. J. Massagué. «Is cancer a disease of self-seeding?». *Nature Medicine* 12 (2006): 875-878.
- Nowell, P. C. «The clonal evolution of tumor cell populations». *Science* 194 (1976): 23-28.
- Olsson, A. K., A. Dimberg, J. Kreuger y L. Claesson-Welsh. «VEGF receptor signalling - in control of vascular function». *Nature Reviews Molecular Cell Biology* 7 (2006): 359-371.
- Padua, D., X. H. Zhang, Q. Wang, C. Nadal, W. L. Gerald, R. R. Gomis y J. Massagué. «TGFbeta primes breast tumors for lung metastasis seeding through angiopoietin-like 4». *Cell* 133 (2008): 66-77.
- Paget, S. «The distribution of secondary growths in cancer of the breast». *Lancet* 1 (1889): 571-573.
- Parato, K. A., D. Senger, P. A. Forsyth y J. C. Bell. «Recent progress in the battle between oncolytic viruses and tumours». *Nature Reviews Cancer* 5 (2005): 965-976.
- Perl, A. K., P. Wilgenbus, U. Dahl, H. Semb y G. Christofori. «A causal role for E-cadherin in the transition from adenoma to carcinoma». *Nature* 392 (1998): 190-193.
- Pouyssegur, J., F. Dayan y N. M. Mazure. «Hypoxia signalling in cancer and approaches to enforce tumour regression». *Nature* 441 (2006): 437-443.
- Roden, R., A. Monie y T. C. Wu. «The impact of preventive HPV vaccination». *Discovery Medicine* 6 (2006): 175-181.
- Rustgi, A. K. «The genetics of hereditary colon cancer». *Genes and Development* 21 (2007): 2.525-2.538.
- Sawyers, C. L. «Targeted cancer therapy». *Nature* 432 (2004): 294-297.
- , «The cancer biomarker problem». *Nature* 452 (2008): 548-552.
- Schiffer, C. A. «BCR-ABL tyrosine kinase inhibitors for chronic myelogenous leukemia». *The New England Journal of Medicine* 357 (2007): 258-265.
- Schmidt-Kittler, O., T. Ragg, A. Daskalakis, M. Granzow, A. Ahr, T. J. Blankenstein, M. Kaufmann, J. Diebold, H. Arnholdt, P. Muller et al. «From latent disseminated cells to overt metastasis: genetic analysis of systemic breast cancer progression». *Proceedings of the National Academy of Sciences USA* 100 (2003): 7.737-7.742.
- Shawver, L. K., D. Slamon y A. Ullrich. «Smart drugs: tyrosine kinase inhibitors in cancer therapy». *Cancer Cell* 1 (2002): 117-123.
- Sweet-Cordero, A., S. Mukherjee, A. Subramanian, H. You, J. J. Roix, C. Ladd-Acosta, J. Mesirov, T. R. Golub y T. Jacks. «An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis». *Nature Genetics* 37 (2005): 48-55.
- Thiery, J. P. «Epithelial-mesenchymal transitions in tumour progression». *Nature Reviews Cancer* 2 (2002): 442-454.
- Van't Veer, L. J., H. Dai, M. J. Van de Vijver, Y. D. He, A. A. Hart, M. Mao, H. L. Peterse, K. Van der Kooy, M. J. Marton, A. T. Witteveen et al. «Gene expression profiling predicts clinical outcome of breast cancer». *Nature* 415 (2002): 530-536.
- Van de Vijver, M. J., Y. D. He, L. J. Van't Veer, H. Dai, A. A. Hart, D. W. Voskuil, G. J. Schreiber, J. L. Peterse, C. Roberts, M. J. Marton et al. «A gene-expression signature as a predictor of survival in breast cancer». *The New England Journal of Medicine* 347 (2002): 1.999-2.009.
- Vogelstein, B., E. R. Fearon, S. R. Hamilton, S. E. Kern, A. C. Preisinger, M. Leppert, Y. Nakamura, R. White, A. M. Smits y J. L. Bos. «Genetic alterations during colorectal-tumor development». *The New England Journal of Medicine* 319 (1988): 525-532.
- , y K. W. Kinzler. «Cancer genes and the pathways they control». *Nature Medicine* 10 (2004): 789-799.
- Vousden, K. H. y D. P. Lane. «p53 in health and disease». *Nature Reviews Molecular Cell Biology* 8 (2007): 275-283.
- Walsh, T. y M. C. King. «Ten genes for inherited breast cancer». *Cancer Cell* 11 (2007): 103-105.
- Wang, C., Y. Yuan y R. H. Hunt. «The association between Helicobacter pylori infection and early gastric cancer: a meta-analysis». *The American Journal of Gastroenterology* 102 (2007): 1.789-1.798.
- Wang, J., R. Loberg y R. S. Taichman. «The pivotal role of CXCL12 (SDF-1)/CXCR4 axis in bone metastasis». *Cancer Metastasis Rev* 25 (2006): 573-587.
- Wang, W. «Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins». *Nature Reviews Genetics* 8 (2007): 735-748.
- Weinberg, R. A. *The biology of cancer*. New York: Garland Science. 2007.
- Welsh, P. L. y M. C. King. «BRCA1 and BRCA2 and the genetics of breast and ovarian cancer». *Human Molecular Genetics* 10 (2001): 705-713.
- Woodman, C. B., S. I. Collins y L. S. Young. «The natural history of cervical HPV infection: unresolved issues». *Nature Reviews Cancer* 7 (2007): 11-22.
- WorldHealthOrganization. Cancer. Fact Sheet No 297. 2008. <http://www.who.int/mediacentre/factsheets/fs297/en/index.html>.
- Yang, J., S. A. Mani, J. L. Donaher, S. Ramaswamy, R. A. Itzykson, C. Come, P. Savagner, I. Gitelman, A. Richardson y R. A. Weinberg. «Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis». *Cell* 117 (2004): 927-939.
- Young, L. S. y A. B. Rickinson. «Epstein-Barr virus: 40 years on». *Nature Reviews Cancer* 4 (2004): 757-768.
- Zur Hausen, H. «Viruses in human cancers». *European Journal of Cancer* 35 (1999): 1174-1181.

el jardín del edén amenazado: ecología y biología de la conservación

CARLOS M. DUARTE

Naciones Unidas ha declarado 2009 como Año Internacional de la Biodiversidad, para homenajear así el bicentenario del nacimiento de Charles Darwin (12 de febrero de 1809-19 de abril de 1882), cuyo libro *Sobre el origen de las especies por la selección natural* (1859) fija el arranque de la ciencia de la biodiversidad. Por fin todo parecía tener sentido: las sutiles diferencias entre especies similares, los coloridos plumajes de las aves y las flores, las numerosas adaptaciones de los animales a su entorno, y los fracasos, reflejados en fósiles de animales fabulosos que, a falta de una explicación, la Iglesia tachó durante décadas de argucias del demonio para confundir a los creyentes. La ciencia había formulado una explicación racional para lo que hasta entonces sólo se podía explicar como el resultado de acciones sobrenaturales. Huelga decir que las revolucionarias tesis de Darwin fueron combatidas con energía durante años. En nuestro país, el biólogo cántabro Augusto G. de L. (Valle de Cabuérniga, Cantabria, 1845-Santander, 1 de mayo de 1904) perdió su cátedra de Historia Natural en la Universidad de Santiago de Compostela por enseñar las teorías de Darwin.

La obra de Darwin, conjugada con las Leyes de Mendel (del monje Gregor J. Mendel, 20 de julio de 1822-6 de enero de 1884) que describen las leyes básicas de la herencia genética, es la semilla de la que germina la bio-

logía moderna, y de ella se desprenden con una imparable secuencia lógica descubrimientos fundamentales como el del ADN y la genómica moderna. Los 150 años que nos separan de la publicación de *El origen de las especies* han estado jalonados de hitos que han conformado una nueva ciencia, la Ecología, que se ocupa de desentrañar las claves del funcionamiento de la biosfera y el papel de la biodiversidad en el equilibrio de la naturaleza, traspasando las fronteras del conocimiento para profundizar, en un momento de crisis aguda, en el pilar sobre el que se asienta el bienestar presente y futuro de la humanidad.

En este capítulo ofrezco un repaso por estos hitos y desarrollos para documentar el camino recorrido para comprender las claves del funcionamiento de la naturaleza y apunto los desafíos que hemos de afrontar durante el siglo XXI. Por vez de seguir un orden cronológico, que ofrecería una visión desordenada del progreso en este campo, he optado por una organización temática en la que señalo los hitos y desafíos más importantes, estructura que supongo más cómoda para el lector.

Origen y diversificación de la vida

El océano es la cuna de la vida sobre la tierra. Los fósiles más antiguos existentes se hallaron en Australia, datados en torno a 3.500 años antes del presente, y correspon-

den a consorcios de microorganismos con Arqueas y cianobacterias fotosintéticas que formaban estructuras de carbonato similares a los estromatolitos que se pueden visitar en distintas zonas del planeta, como en la misma Australia (figura 1).

Los organismos más antiguos existentes son los microorganismos pertenecientes al dominio de las Arqueas que aún representan una fracción importante de las comunidades biológicas del océano profundo. El descubrimiento de las Arqueas es un hito reciente que ha revolucionado nuestra concepción de la organización de la diversidad biológica. En 1977 el microbiólogo estadounidense Carl R. Woese observó, utilizando por primera vez la técnica del RNA ribosómico para establecer las relaciones entre microorganismos, que las comunidades de microorganismos de suelos incluían microorganismos que representaban un nuevo dominio, tan distinto de las bacterias como de los propios eucariotas. El desarrollo de sondas moleculares capaces de discriminar entre bacterias y Arqueas, que son indistinguibles al microscopio, ha permitido detectar que este grupo está presente en todo el planeta, y que son particularmente prominentes en el océano profundo, donde hay hábitats con condiciones semejantes a las existentes cuando las Arqueas aparecieron, pero también en lagos polares. El descubrimiento de las Arqueas llevó a una revisión de los dominios de la vida, para reconocer tres dominios: Bacteria, Archaea y Eukarya, transformando completamente la concepción tradicional.

La atmósfera terrestre primitiva era una atmósfera carente de oxígeno, muy reductora, y carente de ozono, con lo que la radiación ultravioleta penetraba hasta la superficie del planeta Tierra con una intensidad incompatible con la vida. Sólo en el océano, donde la radiación ultravioleta se atenúa fuertemente en profundidad, fue posible que la vida prosperase en este planeta Tierra fuertemente irradiado. La biota marina alteró de forma profunda e irreversible la atmósfera del planeta Tierra y, con ella, las condiciones para la vida en los continentes. En concreto, la aparición de la fotosíntesis oxigénica, por ejemplo, que produce oxígeno por fotólisis del agua (el proceso fotosintético característico de las plantas) en microorganismos marinos, las cianobacterias, produjo un cambio fundamental en la composición de la atmósfera terrestre, con la aparición del oxígeno, que comprende el 21% de la atmósfera terrestre y que al reaccionar con la radiación ultravioleta en la estratosfera (cerca de 12.000 m sobre la superficie terrestre), genera el ozono que absorbe la radiación ultravioleta más dañina y posibilita la vida en los continentes. Disminuyó, además, la concentración de CO₂ en la atmósfera, pues el incremento de O₂ sólo es posible por el consumo proporcional de CO₂ por el proceso de fotosíntesis, almacenado en forma orgánica en el agua oceánica, en suelos, en organismos, detritus y depósitos de petróleo y gas. El cambio de una atmósfera reductora a una atmósfera oxidante es fundamental y condiciona completamente

toda la química planetaria y con ella el funcionamiento de la biosfera y la evolución de la vida.

El origen de las cianobacterias, responsables de este cambio de régimen en el planeta Tierra, que el registro fósil conocido marca como un episodio relativamente abrupto, sigue siendo un misterio y no se puede descartar un origen extraterrestre. De hecho, la aparición de la vida en el océano transformó de forma determinante no sólo la biosfera, el océano y la atmósfera, sino también la litosfera, ya que la formación de carbonato y otros minerales por los organismos marinos es la base de la formación de muchas rocas sedimentarias.

Se conocen fósiles de animales de 800 millones de años de antigüedad, aunque los primeros animales complejos aparecieron hace unos 640 millones de años, de nuevo en Australia. La primera ocupación animal de los continentes se sitúa hace poco más de 400 millones de años durante los que aparecen fósiles de ciempiés y arácnidos. De hecho, la ocupación de los continentes por la vida no hubiera sido posible sin la alteración de las condiciones que ofrecía el planeta Tierra derivada de la actividad de los organismos marinos primitivos.

La historia evolutiva de la vida tiene, por tanto, un recorrido mucho más largo en el océano que en tierra, lo que se refleja en la mayor diversidad de formas de vida presentes en el océano comparadas con las existentes en tierra. Aunque el océano contiene una proporción modesta de las especies que pueblan la Tierra, en él se encuentra el repertorio casi completo de toda la diversidad genómica que la evolución ha generado. La diversidad genómica se refiere a la diversidad de maquinaria genética, integrada por genes que codifican proteínas que intervienen en distintas funciones. Por ejemplo, basta considerar que los genomas de un gusano o de la mosca de vinagre difieren en menos de un 50% de las secuencias del genoma humano, de forma que el recorrido de diversidad genómica de los animales terrestres es relativamente corto.

El árbol de la vida, que refleja la diversificación de las formas de vida en la Tierra, tiene sus raíces en el océano. Así, se conocen treinta filos, las grandes ramas del árbol de la vida, en el mar, trece de ellos exclusivos del océano. En comparación, únicamente se han censado quince filos terrestres, de los que sólo uno es exclusivo de estos ambientes. De hecho, la diversidad de formas de vida en el océano nos causa frecuentemente perplejidad. Por ejemplo, muchos organismos sésiles y coloridos, parecidos a las flores que adornan los paisajes terrestres, son de hecho animales, como anémonas, o mezclas de animales y plantas, como los coloridos colares tropicales, cuyo color se debe a los pigmentos de algas fotosintéticas que habitan los pólipos coloniales que forman el coral. De hecho, las divisiones sencillas entre animal y planta que son útiles en tierra son frecuentemente engañosas en el océano, pues muchos animales son en realidad consorcios de especies fotosintéticas y animales y muchos organismos unicelulares tienen capacidades propias de ambos.

¿Cuántas especies hay en el planeta?

Desde que el científico sueco Carl Linnaeus sentó, con la publicación en 1735 de su obra *Systema Naturae*, los fundamentos de la taxonomía con un sistema de nomenclatura para clasificar los seres vivos, el número de especies descritas ha venido aumentando sin pausa hasta alcanzar cerca de dos millones de especies descritas en la actualidad. El inventario del número de especies en la biosfera parece no tener fin, aunque es evidente que el número de especies existente ha de ser, necesariamente, finito. En los últimos años se han hecho esfuerzos importantes para ofrecer una estima fiable del número de especies que la biosfera puede contener.

A pesar de la extensa historia evolutiva de la vida en el océano, los océanos albergan tan sólo unas 230.000 especies conocidas, casi cincuenta veces menos que el número de especies conocidas en tierra, que alcanza los 1,8 millones (Jaume y Duarte 2006; Bouchet 2006), lo que ha intrigado a los científicos durante muchos años, dando lugar a diversas hipótesis para explicar esta paradoja. Se ha hablado del enorme potencial de dispersión de los propágulos (huevos y larvas) de los animales marinos, que evitaría la segregación genética por la separación de las poblaciones. Por ejemplo, hay sólo 58 especies de plantas superiores, con semillas y frutos (angiospermas) marinos, frente a unas 300.000 en los continentes, y prácticamente no existen insectos en el océano, cuando los artrópodos (que incluyen insectos, crustáceos, arácnidos, ácaros y otros grupos menores) representan el 91% del inventario de especies en tierra.

Se han utilizado distintas aproximaciones para estimar cuál puede ser el número de especies totales, como

extrapolaciones desde los taxones mejor conocidos a los menos conocidos asumiendo una proporcionalidad de especies, extrapolaciones basadas en el número de especies nuevas que aparecen por unidad de área examinada a la superficie total ocupada por distintos hábitats y estimaciones estadísticas basadas en la progresión de la tasa de descubrimiento de nuevas especies. Estas estimas apuntan a que el número total de especies podría situarse en torno a los doce millones, dominadas por insectos con casi diez millones y nemátodos con un millón. El número de especies marinas podría ser ligeramente superior a un millón, algo más de un 10% del número total de especies (Bouchet 2006).

Hallazgos en la exploración de la biodiversidad

Cada año se describen en torno a 16.000 especies nuevas, de las que unas 1.600 son marinas (Bouchet 2006). El crecimiento anual del inventario de biodiversidad se aproxima al 1%. Dado que se piensa que el número de especies descritas en el presente representa cerca de un 10% del total, al ritmo actual de descubrimiento serán necesarios más de doscientos años para completar el inventario, posiblemente más tiempo aún en el caso del inventario de especies marinas, que progresa más lentamente que el inventario de las terrestres. El proyecto internacional Censo de la Vida Marina (www.coml.org) coordina los esfuerzos de miles de investigadores en todo el mundo con el objetivo de llegar a un inventario de todas las especies existentes en el océano. Cada año se describen 1.635 nuevas especies marinas, en su gran mayoría crustáceos y moluscos por parte de 2.000 taxónomos marinos en activo (Bouchet 2006). Sin embargo, se ha estimado que a este ritmo de descubrimiento necesitaremos de 250 a 1.000 años para finalizar el inventario de biodiversidad marina, que podría situarse en torno al millón y medio de especies, seis veces las descritas actualmente (Bouchet 2006).

Esta labor de inventario reporta importantes sorpresas, no sólo en organismos microscópicos, sino en vertebrados relativamente grandes como monos (el mono mangabey *Lophocebus kipunji* descubierto en Tanzania en 2005) o peces. Aunque el número de especies que se descubre cada año en tierra supera de largo al de especies marinas, los descubrimientos en tierra se limitan a nuevas especies de géneros o familias conocidas, mientras que el rango taxonómico de innovaciones en el océano es muy superior. El conocimiento de la diversidad de la vida en el océano es aún muy limitado, y la tasa de descubrimientos es aún sorprendentemente elevada.

Algunos de los descubrimientos dentro de los organismos microscópicos suponen, también, importantes hitos en nuestro conocimiento. Por ejemplo, las diminutas cianobacterias fotosintéticas de los géneros *Synechococcus* (aproximadamente 1 μm de diámetro) y *Prochlorococcus* (aproximadamente 0,5 μm de diámetro) fueron descubiertas a finales de los años 1970. Estudios posteriores



Figura 1. Estromatolitos en Shark Bay, Western Australia, la localidad donde se descubrieron por primera vez estromatolitos vivos (fotografía del autor).

pusieron de manifiesto que estos organismos dominan el plancton de los grandes desiertos oceánicos, que representan casi un 70% de la extensión del océano abierto y son responsables de casi un 30% de la producción fotosintética en el océano. La magnitud de este descubrimiento y lo que nos dice sobre el grado de desconocimiento del océano se puede comprender considerando que el desconocimiento de estos organismos hasta finales de los años 1970 equivale a que hubiésemos desconocido la existencia de las selvas tropicales en tierra hasta esa fecha. El océano sigue asombrándonos a niveles taxonómicos elevados, incluso se descubren nuevos filos —cosa que no ocurre en tierra—, y algunos de los animales más grandes del planeta, como los calamares gigantes *Magnapinnidae* de enormes aletas, que se han avistado varias veces en el océano profundo (> 2,000 m), el tiburón de boca ancha, *Megachasma pelagios* —de 4,5 m de longitud— descubierto en aguas del Indopacífico en 1983, o el pequeño rorcual *Balaenoptera omurai* que alcanza los 9 m de longitud —descubierto en la misma zona en 2003.

Las mayores oportunidades de lograr hallazgos en biodiversidad marina se dan en los hábitats remotos o extremos. En tierra la mayor parte de los descubrimientos más espectaculares provienen frecuentemente de selvas tropicales en lugares remotos y relativamente inexplorados de Asia (Vietnam), África (Tanzania) y Oceanía (Papua Nueva Guinea). En el océano, son las zonas remotas del Sureste Asiático y Oceanía, centros de gran diversidad en especies de todos los grupos marinos, y los hábitats extremos —como fosas oceánicas, cuevas submarinas, ambientes hipersalinos y anóxicos, fuentes hidrotermales, y bolsas de aguas hipersalinas y anóxicas— las que reportan mayores sorpresas (Duarte 2006), junto con el interior de organismos que albergan simbioses, un término que engloba a comensales, mutualistas y parásitos, y que no están restringidos a especies de pequeño tamaño. Por ejemplo, el mayor gusano marino conocido, que alcanza 6 m de longitud, es un parásito de ballenas.

Los hallazgos sobre biodiversidad marina en el océano van mucho más allá de la descripción de nuevas especies, por sorprendentes que éstas sean, para incluir hallazgos de ecosistemas con comunidades y procesos metabólicos desconocidos hasta la fecha. A finales de los años 1970 los científicos a bordo del sumergible norteamericano de investigación *Alvin* descubrieron los ecosistemas de fuentes hidrotermales mientras realizaban estudios geotérmicos en la dorsal de las Galápagos (Lonsdale 1977; Corliss et al. 1979). Encontraron un extraordinario paisaje de chimeneas negras de las que emanaba un fluido con aspecto de humo, compuesto de metales y otros materiales que al enfriarse precipitaban haciendo crecer estas chimeneas, que estaban colonizadas por densas masas de animales hasta entonces desconocidos, como el gusano tubícola gigante *Riftia pachyptila*, cangrejos albinos, peces y otros muchos organismos, todos ellos nuevos para la ciencia.

Este descubrimiento no sólo supuso un importante impulso al inventario de especies marinas, sino un desafío completo a nuestra creencia de que la luz solar era la fuente de energía que, a partir de la actividad fotosintética de las plantas, permitía la producción de materia orgánica que mantiene los ecosistemas. En los arrecifes llenos de vida de las fuentes hidrotermales no eran las plantas quienes transformaban la energía en la materia orgánica que alimentaba el ecosistema, sino que este trabajo lo hacían bacterias y Arqueas quimioautotróficas que sintetizan materia orgánica a partir de compuestos inorgánicos reducidos que fluyen del interior de la Tierra en los fluidos hidrotermales (Karl, Wirsén y Jannasch 1980; Jannasch y Mottl 1985). Estos nuevos hábitats donde la vida prospera sin necesidad de energía solar son conocidos como ecosistemas quimiosintéticos, en los que los microorganismos establecen relaciones simbióticas con los invertebrados. Desde su descubrimiento, en 1977, se han descrito cerca de 600 especies de organismos que los habitan. Desde entonces se ha descubierto que otros hábitats reductores del lecho marino, como las surgencias frías de fluidos hidrotermales (descubiertas en 1983 a 500 m en el Golfo de México), los restos de ballenas y zonas con mínimos de oxígeno, también albergan comunidades que dependen de la energía química, parecidas a las de los animales de las fuentes hidrotermales.

Estos descubrimientos suponen un hito revolucionario en cuanto que modifican completamente nuestras ideas sobre el funcionamiento y organización de ecosistemas. Los microorganismos encontrados en las fuentes hidrotermales han supuesto también una pequeña revolución de la biología y la biotecnología, pues muchos de ellos tienen proteínas que son estables a 100°C de temperatura y que catalizan reacciones a una velocidad vertiginosa. *Pyrococcus furiosus* es una especie de Archaea descubierta en fondos marinos de la isla de Vulcano (Italia) que se destaca por tener una temperatura de crecimiento óptimo de 100°C que le permite duplicarse cada 37 minutos y posee enzimas que contienen tungsteno, un elemento que rara vez se encuentra en las moléculas biológicas. A esta temperatura las polimerasas del ADN de *Pyrococcus furiosus* (Pfu ADN) operan a una enorme velocidad, por lo que se utilizan a menudo en la reacción en cadena de la polimerasa (PCR), que permite producir masivamente copias de fragmentos de ADN y es el fundamento de la mayoría de aplicaciones biotecnológicas que requieren secuenciación de ADN.

Los nuevos descubrimientos en biodiversidad marina también dependen de desarrollos en técnicas moleculares que permiten establecer la posición taxonómica de organismos a partir del análisis de secciones de su genoma. Por ejemplo, el uso de técnicas de secuenciación masiva permitió al biólogo americano Craig Venter, que lideró el proyecto de Celera Genomics que secuenció por primera vez el genoma humano, secuenciar fragmentos de ADN de

1 m³ de agua superficial del Mar de los Sargazos. Este ejercicio arrojó un sorprendente inventario de nuevos genes, 1.214.207, y cerca de 1.800 nuevas especies de microorganismos microbianos (Venter et al. 2004). Lamentablemente estas técnicas no permiten averiguar la identidad de esas nuevas especies, pero sí están revelando que muchas especies marinas anatómicamente similares resultan ser especies distintas. Por otro lado, estas técnicas también están demostrando que algunas especies consideradas distintas en base a sus diferencias morfológicas resultan ser, en realidad, variantes de una misma especie sometidas a condiciones ambientales muy distintas.

La biosfera bajo presión: el Antropoceno

La Revolución Industrial, con el aumento de la capacidad del ser humano de transformar su entorno, supone un hito no ya en la historia de nuestra especie, sino en la del planeta, que se ha visto transformado por la actividad humana. Cualquier examen objetivo del planeta Tierra, su clima, la configuración y dinámica de sus ecosistemas, sus procesos funcionales básicos, muestra que éstos se encuentran afectados por la actividad humana. La importancia del impacto de la actividad humana sobre los procesos esenciales de la biosfera se refleja en algunos indicadores como el hecho de que el 45% de la superficie terrestre se ha transformado ya por la actividad humana, pasando de ecosistemas salvajes a ecosistemas domesticados, como campos de cultivo, pastizales o zonas urbanas. La humanidad utiliza más de la mitad del flujo disponible de agua dulce global, modificando no sólo la cantidad de agua que fluye por ríos sino también alterando su calidad, que resulta enriquecida en nutrientes, nitrógeno y fósforo, materia orgánica y contaminantes tras su uso por la humanidad. De hecho, la actividad humana acelera de forma notable los ciclos de los elementos en la biosfera, habiendo movilizado más de 420 Gigatoneladas de carbono desde la Revolución Industrial y fijando, a través de la reacción de Haber, patentada en 1908 por Fritz Haber, 154 Megatoneladas anuales de nitrógeno, gas atmosférico en forma de amonio para su uso en fertilizantes, más nitrógeno atmosférico que los procesos de fijación de nitrógeno que ocurren a través de la actividad nitrogenosa presente en plantas y microorganismos terrestres y marinos. Las emisiones de dióxido de carbono por el uso de combustibles fósiles, producción de cemento y fuegos, junto con la liberación de otros gases con efecto invernadero como el metano, están incrementando la temperatura del planeta y, al disolverse en el océano, acidificando el océano, procesos que tienen consecuencias importantes para el clima de la Tierra y sobre los ecosistemas que contiene. Se ha calculado también que la humanidad se apropia, a través de la explotación agrícola, forestal y pesquera, de aproximadamente 40% de toda la producción fotosintética terrestre y el 20% de la producción fotosintética costera del planeta.

Estos datos, a los que se podrían añadir muchos más, bastan para sustanciar la afirmación de que nuestra especie se ha erigido en un motor esencial de cambio en los procesos básicos de la biosfera, lo que llevó al químico atmosférico y premio Nobel Paul Crutzen a proponer, junto con su colega E. Stoermer, en el año 2000 el nombre *Antropoceno* para designar una nueva Era geológica en la historia del planeta en la que la humanidad ha emergido como una nueva fuerza capaz de controlar los procesos fundamentales de la biosfera (Crutzen y Stoermer 2000), causando un Cambio Global.

La capacidad de la humanidad para alterar el planeta se inicia con el Holoceno, al finalizar el último período glacial hace cerca de 10.000 años, que fue seguido por el desarrollo y rápida expansión de la agricultura, la ganadería y los primeros núcleos urbanos. Los primeros indicios de esta nueva fuerza emergente son las extinciones de grandes mamíferos y aves, cazados por los primeros pobladores de islas y continentes. El desarrollo de la agricultura y la ganadería condujo a la transformación del territorio, transformando bosques y otros ecosistemas en campos de cultivo y pastizales, cambios favorecidos por la capacidad de trabajo que generaron la domesticación de animales de trabajo (bueyes, caballos, etc.) y desarrollos tecnológicos como el arado y la rueda. La capacidad de la humanidad para transformar el planeta experimentó un notable impulso con la Revolución Industrial, que aumentó la capacidad de utilizar energía para transformar el planeta, generando a la vez residuos, como gases y compuestos sintéticos, que alteran los procesos naturales. La humanidad ha transformado el territorio del planeta de forma radical, transformando cerca del 45% de la superficie terrestre en pastizales, que ocupan en torno al 30% de la superficie terrestre, campos de cultivo, que ocupan cerca del 10% de la superficie, y zonas urbanas que ocupan aproximadamente el 2% de la superficie terrestre. Otras infraestructuras, como embalses, carreteras, tendidos eléctricos, ferroviarios, etc., ocupan en torno a un 3% adicional de la superficie del planeta. Las zonas costeras están experimentando las tasas de crecimiento demográfico más altas del planeta. Cerca del 40% de la población humana vive a menos de 100 kilómetros de la costa, con una densidad de población tres veces superior a la de territorios continentales, y la población costera está creciendo mucho más rápido que la continental debido a la migración, aumento de la fertilidad en las zonas costeras y aumento de los flujos turísticos a estas áreas (Millennium Assessment 2005b), y la línea de costa está siendo rápidamente ocupada por infraestructuras (urbanizaciones, calles y carreteras, puertos, etc.).

La actividad humana ha acelerado los ciclos de los elementos en la biosfera, que constituyen uno de los procesos centrales en la regulación del funcionamiento de este sistema y de la vida. La aceleración de los ciclos elementales afecta prácticamente a todos los elementos quími-

cos, pero sus consecuencias son más importantes sobre aquellos que participan de procesos esenciales en la regulación de procesos vitales, como son el carbono, nitrógeno, fósforo, hierro, calcio y otros oligoelementos, y los que participan en procesos de regulación climática, como es el carbono, a través del CO_2 y el metano, y el nitrógeno, a través del óxido nitroso. La transformación del territorio desde bosques a pastizales y campos de cultivo supone una aceleración del ciclo de carbono, que pasa de estar atrapado en biomasa de bosques, a reciclarse rápidamente en cosechas anuales. Los suelos agrícolas tienen menor capacidad de retención de carbono que los suelos forestales, y la destrucción de humedales ha puesto en circulación el carbono retenido en estos sistemas, que son sumideros importantes de carbono. La extracción de combustibles fósiles y gases supone igualmente una movilización de carbono acumulado durante épocas en las que la biosfera generaba un exceso de producción primaria.

El uso de combustibles fósiles, junto con la producción de CO_2 durante la producción de cemento, la deforestación y los fuegos forestales han emitido alrededor de 450 Gt de CO_2 a la atmósfera, lo que ha llevado a un rápido aumento de la concentración de CO_2 , acompañada de la de otros gases con efecto invernadero, como el metano y el óxido nitroso, en la atmósfera. La actividad humana genera también un exceso de movilización de nitrógeno, fundamentalmente a partir de la producción de unos 154 millones de toneladas anualmente en forma de fertilizantes a partir de nitrógeno gas atmosférico. Este nitrógeno se moviliza a través de su transporte en ríos, transporte atmosféricos y también por la contaminación por nitratos de los acuíferos. El transporte atmosférico permite el transporte a larga distancia del nitrógeno, que se deposita también en el océano abierto. La producción de fertilizantes requiere de la extracción, a partir de depósitos mineros, de una cantidad de fósforo proporcional a la cantidad de nitrógeno producida en fertilizantes. La aceleración de los ciclos de los elementos tiene importantes consecuencias sobre los ecosistemas, que se ven alterados a través de un proceso conocido como eutrofización, causado por el aporte excesivo de nutrientes a los ecosistemas, y que tiene consecuencias importantes sobre éstos.

La humanidad usa actualmente el 50% del agua dulce disponible en la biosfera, extrayendo, en 1995, más de 3.000 km^3 de agua para irrigación de campos de cultivos. La producción de alimento, incluidos los pastizales, consume anualmente en torno a 14.000 km^3 de agua. Como consecuencia de este uso de agua para la agricultura, grandes lagos, como el Mar de Aral, en Asia Central, han perdido gran parte de su extensión y volumen de agua, reduciéndose su nivel en 0,6 m cada año, y la superficie ocupada por el lago Chad, en África, se redujo en veinte veces en tan sólo quince años. El uso del agua por la humanidad y la transformación del territorio han resultado en importantes cambios en el ciclo del agua.

Aproximadamente el 60% de las zonas húmedas europeas existentes en 1800 se han perdido. El número de embalses construidos ha crecido rápidamente durante el siglo xx, a un ritmo de un 1% anual, reteniendo un volumen de agua de aproximadamente 10.000 km^3 , equivalente a cinco veces el volumen de agua contenida en los ríos.

La actividad humana ha sintetizado millones de nuevos compuestos químicos inexistentes en la biosfera que se comportan, en muchos casos, como contaminantes dañinos para los organismos, incluida nuestra propia especie, o presentan actividades que interfieren con otros procesos, como pueden ser los freones y halones, gases utilizados en procesos industriales y refrigeración, responsables de la destrucción de la capa de ozono, que decayó aproximadamente un 4% anual durante las últimas dos décadas causando la ampliación del agujero de ozono sobre el hemisferio Sur. Mientras estos compuestos han sido ya controlados, a través del Protocolo de Montreal de 1987, cada año siguen liberándose a la biosfera miles de nuevas sustancias sin que existan pruebas previas que indiquen cuáles pueden ser sus impactos sobre la salud humana y la biosfera. Algunos de estos gases se comportan como gases de efecto invernadero y agravan el proceso de calentamiento global. Muchos de estos compuestos son volátiles o semivolátiles, transportándose por la atmósfera hasta lugares alejados miles de kilómetros de sus fuentes, de forma que no existe lugar en la biosfera al que no hayan llegado estos compuestos.

Las emisiones de gases de efecto invernadero están causando un fuerte aumento de la temperatura del planeta, que se ha calentado ya más de 0,7°C y se prevé que pueda suponer un calentamiento de entre 2 y 5 grados adicionales durante el siglo xxi (Trenberth et al. 2007, Meehl et al. 2007). Además del incremento de la temperatura otros componentes del sistema climático se verán igualmente afectados. Así se esperan cambios importantes en el régimen hídrico, con un aumento de la precipitación en algunas zonas del planeta, así como una disminución de la precipitación en otras y eventos extremos (sequías e inundaciones) más frecuentes y prolongados (Meehl et al. 2007). La intensidad del viento aumentará y se espera que eventos extremos como ciclones tropicales aumenten en intensidad y se extiendan a áreas hasta ahora libres de este tipo de fenómenos (Meehl et al. 2007).

El calentamiento global ha supuesto un aumento promedio de unos 15 cm durante el siglo xx, y un aumento adicional de entre 30 y 80 cm proyectado para el siglo xxi (Bindoff et al. 2007). El aumento de la presión parcial de CO_2 en la atmósfera y su penetración en el océano llevan a la disminución del pH del océano, que se puede cuantificar en aproximadamente 0,15 unidades que, dado que la escala de pH es logarítmica, supone un incremento de la acidez del océano en un 60%. El aumento de la presión parcial de CO_2 previsto para el siglo xxi llevará a una disminución adicional de entre 0,3 y 0,4 unidades, lo que supone que para entonces la acidez del océano se habrá triplicado.

Impacto del Cambio Global sobre los ecosistemas

La transformación del territorio por la expansión de los pastizales, campos de cultivo y zonas urbanas e industriales se ha llevado a cabo a costa de la pérdida de ecosistemas, como humedales —muchos de los cuales han sido drenados—, bosques tropicales y otros hábitats críticos para la conservación de la biodiversidad. Los humedales representan un 6% de la superficie de la Tierra, habiéndose perdido más del 50% del área de humedales en Norteamérica, Europa, Australia y Nueva Zelanda, con una extensión importante degradada en estas y otras regiones. En la cuenca Mediterránea se perdieron más del 28% de los humedales durante el siglo xx. Los bosques han sufrido pérdidas importantes que han supuesto la desaparición de aproximadamente el 40% del área forestal del planeta en los últimos tres siglos. Los bosques han desaparecido por completo de 25 países, y otros 29 han perdido más del 90% de su cubierta forestal. Aunque las áreas forestales se están expandiendo actualmente en Europa y Norteamérica, continúan perdiéndose en los trópicos a un ritmo de 10 millones de hectáreas por año, o aproximadamente un 0,5% por año (Millenium Assessment 2005b). La intensa ocupación de la zona costera está causando importantes pérdidas de ecosistemas costeros, que están experimentando las mayores tasas de pérdida: aproximadamente un 35% del área de manglares se ha perdido, cerca de una tercera parte de los arrecifes de corales han sido destruidos (Millenium Assessment 2005b), y las praderas submarinas se están perdiendo a un ritmo en torno al 2–5% anual (Duarte 2002).

El calentamiento del planeta está llevando a cambios espectaculares en el área ocupada por superficies heladas en nuestro planeta, tales como la extensión del hielo marítimo en el Ártico, que sufrió una disminución catastrófica en 2007, y la extensión de glaciares alpinos en clara recesión por el calentamiento del planeta.

El aumento de la presión parcial de CO₂ incrementará las tasas de fotosíntesis, sobre todo de organismos fotosintéticos acuáticos, ya que la enzima responsable de la fijación de CO₂ evolucionó cuando la concentración era mucho mayor que la actual, y su actividad resulta relativamente ineficiente a las concentraciones actuales de CO₂. El aumento de la temperatura también contribuye al aumento de la actividad fotosintética, al acelerar todos los procesos metabólicos. Sin embargo, la respiración es un proceso mucho más sensible al aumento de temperatura y se prevé que la respiración en la biosfera, que está dominada por procesos microbianos, podría incrementarse hasta un 40% con el calentamiento previsto, mientras que la producción primaria aumentaría en torno a un 20% (Harris et al. 2006). Esto podría llevar a una producción neta de CO₂ en los ecosistemas acuáticos, incluido el océano, que vendría a agravar el efecto invernadero.

El proceso de eutrofización, que resulta de la movilización de grandes cantidades de nitrógeno y fósforo por

la actividad humana, está llevando a un incremento de la producción primaria terrestre y acuática. La eutrofización conlleva el deterioro de la calidad del agua, la pérdida de vegetación sumergida, el desarrollo de proliferaciones algales, algunas de ellas tóxicas, y, cuando concurren otras circunstancias, como una pobre ventilación de las aguas, la proliferación de fenómenos de hipoxia. El fenómeno de eutrofización no sólo se limita a los continentes sino que además afecta al océano abierto, para el que los aportes de nitrógeno por vía atmosférica se han duplicado, con consecuencias sin duda importantes, pero pobremente establecidas aún, sobre el funcionamiento de los océanos.

Las huellas del cambio climático son particularmente evidentes en los patrones fenológicos de los organismos. Los patrones de comportamiento y reproducción también están sufriendo y sufrirán alteraciones, con floraciones más tempranas en latitudes templadas y alteraciones en los periodos de migración de aves. Las actividades que los organismos inician en primavera en latitudes templadas se están anticipando. El aumento de temperatura está causando ya cambios en los rangos biogeográficos de los organismos, desplazándose hacia latitudes más altas. Estos desplazamientos incluyen a organismos patógenos, con lo que se espera que el rango de enfermedades de carácter tropical o subtropical se extienda a latitudes más altas. Además de desplazamientos latitudinales, estos desplazamientos también se extienden a cambios en el rango de elevación en los que ocurren distintos organismos. El límite de los bosques en alta montaña se está ampliando a elevaciones más altas y los organismos alpinos vienen aumentando su límite superior entre 1–4 m por década. Como consecuencia de esto se están dando cambios relativamente importantes en la composición de las comunidades prácticamente en todos los ecosistemas del planeta.

El Cambio Global parece conducir, a partir de la conjunción de múltiples efectos (calentamiento y eutrofización), a un aumento de la hipoxia como un problema en aguas costeras, con un incremento de un 5% anual en las áreas que experimentan hipoxia (Vaquer-Sunyer y Duarte 2008). La hipoxia supone la caída, temporal o permanente, de la concentración de oxígeno de las aguas costeras por debajo de entre 2 y 4 mg/L de oxígeno, lo que conlleva la mortalidad de muchos grupos de animales y plantas y la liberación de fósforo del sedimento. El fenómeno de hipoxia requiere el concurso de tres circunstancias: (a) un exceso de producción fotosintética que sedimenta a las aguas en contacto con el fondo marino; (b) la estratificación, mediante un gradiente de densidad debido a un gradiente térmico, un gradiente de salinidad o ambos, entre las aguas superficiales en contacto con la atmósfera y las aguas costeras más profundas en contacto con el sedimento marino, de forma que esta estratificación supone una barrera a la ventilación del agua y la renovación de su contenido en oxígeno; y (c) el aumento de la respiración en la capa de agua profunda. Estos tres

procesos se encuentran afectados por el proceso de Cambio Global: el proceso de eutrofización global está incrementando la producción costera, a partir del incremento de los aportes de nitrógeno y fósforo; el aumento de temperatura aumentaría la estratificación de la columna de agua, reduciendo la ventilación de las aguas subyacentes, e incrementaría la tasa de respiración. Así se prevé que el Cambio Global aumente notablemente la extensión e intensidad de los problemas de hipoxia, y la mortalidad de organismos marinos que ésta ocasiona, en el océano costero (Vaquer-Sunyer y Duarte 2008).

La acidificación del océano afecta, principalmente, a los organismos que tienen esqueletos carbonatados. Los organismos de océanos fríos son particularmente vulnerables a este proceso, de forma que es en los océanos polares donde la acidificación causará antes problemas en el desarrollo de organismos con estructuras calcificadas, dificultades que más adelante afectarán también a los organismos de mares templados y eventualmente a los organismos tropicales.

Los arrecifes de coral son particularmente vulnerables al aumento de la temperatura, pues los simbioses fotosintéticos que contienen, y de los que dependen para un adecuado crecimiento, mueren cuando la temperatura del agua supera los 29°C, lo que ocurrirá con mayor frecuencia en el futuro. De hecho los arrecifes de coral del Sureste Asiático han experimentado recientemente episodios masivos de blanqueamiento (pérdida de zooxantelas simbioses). Se teme que los arrecifes de coral, que además sufren las consecuencias de la eutrofización global y la acidificación del agua de mar, se encuentren entre los ecosistemas más impactados por el Cambio Global.

Finalmente, la pérdida acelerada de superficie de hielo en verano en el Ártico está colocando en serio peligro a las especies, como osos polares, focas y morsas, que dependen del hielo como hábitat.

La respuesta de los ecosistemas a estas presiones simultáneas conlleva, frecuentemente, cambios abruptos en las comunidades, que se conocen como cambios de régimen y que suponen transiciones bruscas entre dos estados (lagos someros dominados por vegetación enraizada en el fondo a lagos dominados por fitoplancton con la eutrofización; fondos marinos con vegetación y fauna a fondos marinos dominados por tapetes microbianos con la hipoxia). Estas transiciones se producen tras un pequeño incremento en las presiones suficientes para cruzar un umbral que dispara el cambio. Las primeras especulaciones teóricas sobre estos saltos bruscos de régimen en el estado de ecosistemas data de los años 1970 (May 1977). Desde entonces se ha constatado que estos cambios no son la excepción, sino la respuesta más frecuente de ecosistemas sometidos a presiones (Scheffer y Carpenter 2003, Andersen et al. 2008). También se ha constatado que una vez cruzados estos umbrales que disparan el cambio de régimen es muy difícil revertir el sistema al estado

anterior. De ahí la importancia de determinar la posición de estos umbrales de presiones. Lamentablemente, hoy en día sólo somos capaces de identificar estos umbrales una vez que los hemos cruzado (Strange 2008).

¿Hacia la sexta extinción?:

extinciones y la crisis de biodiversidad

El que las especies se extingan es tan natural como el que surjan como resultado del lento proceso evolutivo. El registro fósil aporta evidencias de que ha habido cinco grandes extinciones en el turbulento pasado de nuestro planeta: la primera hace cerca de 440 millones de años, aparentemente y debido a un cambio en el clima que supuso la pérdida del 25% de las familias existentes. La segunda gran extinción, con la pérdida del 19% de las especies, ocurrió hace 370 millones de años, posiblemente debido a un cambio en el clima global. La tercera gran extinción, la mayor de ellas, tuvo lugar hace 245 millones de años posiblemente por un cambio climático causado por el impacto de un gran meteorito, y supuso la pérdida de 54% de las familias existentes. La cuarta gran extinción, hace 210 millones de años, causó la pérdida de 23% de las familias, y sus causas son motivo de especulación, incluyendo un posible incremento de la radiación ultravioleta debido a una supernova. La quinta, y más famosa de las grandes extinciones tuvo lugar hace 65 millones de años, provocada por el impacto de un gran meteorito seguido de una cadena de grandes erupciones volcánicas que causaron la pérdida del 17% de las familias existentes, incluidos los dinosaurios.

La base de datos de la Unión Internacional para la Conservación de la Naturaleza (IUCN, www.iucnredlist.org) reporta 850 especies actuales ya extintas, la mayor parte de ellas (583 especies) terrestres y de agua dulce (228 especies), con tan sólo 16 especies marinas extinguidas. El número de especies clasificadas por la IUCN en estado crítico es de 3.124 y otras 4.564 especies se encuentran en peligro de extinción.

Estas estimas del número de especies en peligro son conservativas pues necesariamente sólo pueden considerar las especies conocidas, que suponen aproximadamente una de cada diez especies existentes, y porque para que una especie se considere extinta han de transcurrir más de diez años de búsqueda intensa sin avistamientos tras la última vez en la que el organismo fue observado, con lo que pueden haber especies extintas hace varios años que aún no se hayan consignado como tales. Cada año se documenta la desaparición de cerca de doscientas especies en todo el mundo, aunque se estima que este número pueda ser muy superior si se tienen en cuenta las extinciones de especies antes de que sean descritas. Algunas autoridades, incluido el biólogo E. O. Wilson, padre de la biología de la conservación, consideran que cada año se extinguen varias decenas de miles de especies. Esto significa que, al ritmo actual, a finales del siglo **xxi** habrá

desaparecido entre una tercera parte y la mitad del total de especies de la Tierra. Es indudable que estamos viviendo —y causando— una grave crisis de biodiversidad (Eldredge 1998). La extinción de especies por la actividad humana no es, sin embargo, un fenómeno reciente, pues los registros fósiles contienen información abundante sobre muchas especies, en particular grandes mamíferos y aves, que se extinguieron tras la llegada del hombre, particularmente en América y Australia, así como la extinción de fauna del Pleistoceno por la caza.

La transformación del territorio es uno de los principales motores de extinciones, al causar una enorme pérdida de hábitat que ha llevado a la extinción de muchas especies. La pérdida de humedales, en particular, ha tenido un efecto devastador sobre las numerosas especies de árboles, plantas, aves, peces, anfibios e insectos que habitan en estas zonas. Muchas de las extinciones contemporáneas afectan a especies en ambientes insulares, donde los procesos de especiación habían sido particularmente importantes, dando lugar a un elevado número de endemismos que resultan más vulnerables a la actuación humana. Las introducciones de especies con comportamiento invasor por la actividad humana ha sido también un proceso importante en la pérdida de especies. Así, la introducción del zorro y el gato en el continente australiano diezmo los pequeños marsupiales, muchos de ellos ya extintos y otros en grave peligro de extinción. Las especies invasoras afectan a la biodiversidad local, desplazando especies autóctonas; su comportamiento agresivo se explica frecuentemente por la ausencia de predadores y parásitos en las nuevas áreas donde se han introducido. La actividad humana ha introducido, por ejemplo, más de 2.000 especies de plantas en Estados Unidos y Australia, y unas 800 en Europa (Vitousek et al. 2003). En algunos casos las especies invasoras pueden tener efectos positivos sobre el ecosistema, así, por ejemplo, la presencia del mejillón cebra, que invade ríos y estuarios en Europa y Norteamérica, puede atenuar los efectos de la eutrofización sobre estos ecosistemas.

La actividad humana también ha afectado de forma importante la biodiversidad marina. En particular, la sobrepesca ha reducido la biomasa de peces en el océano a la décima parte de la existente a principios del siglo xx (Millenium Assessment 2005). La creciente presión sobre los ecosistemas costeros está derivando en una crisis de biodiversidad de dimensiones globales reflejada en una pérdida de hábitats de gran valor ecológico (arrecifes de coral, marismas, manglares y praderas submarinas), junto con la biodiversidad que albergan.

Los análisis disponibles apuntan a que un aumento de la temperatura de más de 2°C causaría extinciones de anfibios y corales, y que un aumento de la temperatura por encima de 4°C, dentro de las predicciones de las proyecciones climáticas para este siglo, podría causar una mortalidad masiva, que podría afectar a una de cada tres

especies (Fischlin et al. 2007), comparable a las grandes extinciones del pasado. Un análisis reciente (Mayhew et al. 2007) ha comparado la tasa de extinciones con el cambio de temperatura promedio global, mostrando cómo existe una correspondencia entre cuatro de las cinco grandes extinciones del pasado y cambios climáticos. Esta correspondencia refuerza las predicciones que apuntan a que el cambio climático actual podría causar una nueva extinción masiva (Thomas 2004).

La acción sinérgica de las distintas fuerzas responsables del Cambio Global es la fuerza que impulsa la notable erosión de la biodiversidad. Por ejemplo, los anfibios parecen estar en declive a escala global por causas aún mal definidas pero que parecen relacionadas con una conjunción de causas: la pérdida de hábitat, la lluvia ácida, la contaminación ambiental, el incremento de radiación ultravioleta y el cambio climático. De hecho, el ritmo actual de extinciones de especies ha alcanzado niveles suficientemente altos como para que se pueda pensar que algunos investigadores hayan postulado que hemos entrado ya en la sexta gran extinción.

Ecología y biología de la conservación: claves de nuestro futuro

La constatación de la pérdida de especies y ecosistemas a escalas que van desde fenómenos locales a globales ha impulsado una intensa actividad de investigación en los últimos veinte años para evaluar las consecuencias de las extinciones, el papel de la biodiversidad en el funcionamiento de los ecosistemas y, eventualmente, los beneficios que la biodiversidad reporta a la sociedad. A la vez, un mayor conocimiento de la biología de las especies ha permitido mejorar las posibilidades de conservarlas. Nació la ecología y la biología de la conservación.

A través de experimentos a gran escala se ha podido verificar que, en general, una mayor biodiversidad se corresponde con una mayor producción biológica, un reciclado de nutrientes más eficiente y una mayor resistencia de los ecosistemas frente a perturbaciones (Schwartz et al. 2000). Se han valorado los bienes y servicios que los ecosistemas reportan a la sociedad, estimándose el valor añadido de estos bienes y servicios (provisión de alimento, depuración de aguas, reciclado de nutrientes, regulación de los gases atmosféricos y el clima, polinización, control de patógenos y sus vectores, etc.) en más del doble del producto bruto de todas las naciones (Costanza et al. 1988). La pérdida de estas funciones por el deterioro de los ecosistemas y pérdida de biodiversidad supone una reducción del capital natural que conlleva un grave impacto económico y la pérdida de nuestra calidad de vida.

La Convención para la Diversidad Biológica (www.cbd.int), firmada en Río de Janeiro en 1992 por la mayor parte de las naciones, con notables excepciones, supone una reacción frente a esta crisis de biodiversidad. La Convención parte del reconocimiento del valor intrínseco de la

biodiversidad y su importancia para mantener los sistemas de soporte vitales de los que depende la sociedad, y de la evidencia de que la biodiversidad está siendo erosionada por la actividad humana. La Convención persigue asegurar la conservación de la biodiversidad en el planeta y la distribución equitativa de la riqueza que su uso genera. Uno de sus objetivos es llegar a proteger un 10% de la superficie del planeta. Con este impulso, las áreas protegidas han proliferado y, aunque en tierra se acercan poco a poco a este objetivo, en el océano aún permanecen muy lejos de él.

La protección del territorio se complementa con medidas de protección especial de especies en peligro. Muchas de éstas son especies carismáticas para cuya conservación no se escatiman esfuerzos a través de planes de reproducción cada vez más sofisticados y costosos, que incluyen ya la consideración de la aplicación de avances en técnicas de clonación para su conservación. La clonación es una técnica que se desarrolló en primer lugar usando anfibios y que podría contribuir a la conservación de estos organismos, que se encuentran en una grave situación de riesgo de extinción (Holt et al. 2004). Una iniciativa reciente ha sido la inauguración, en una isla noruega del Ártico, de la Bóveda Global de Semilla de Svalvard, un banco mundial que preserva semillas de especies de interés agrícola de todo el mundo frente a posibles catástrofes (véase www.nordgen.org/sgsv/). Tanto esta infraestructura como el riesgo al que atiende pertenecían, hasta hace poco, al ámbito de la ciencia ficción catastrofista.

El ritmo de extinciones y pérdida de ecosistemas progresa imparable, a pesar de los avances en la protección de áreas naturales y la conservación de especies concre-

tas. Es cada vez más claro que las áreas protegidas o los esfuerzos de conservación de especies individuales sólo pueden entenderse como soluciones parciales frente a los impactos responsables de la pérdida de ecosistemas y biodiversidad, y que han de completarse con otras estrategias y técnicas. Es necesario comprender mejor las causas de las pérdidas de especies, la relación entre las distintas presiones que las provocan, la posibilidad de efectos dominó en las extinciones de especies (Rezende et al. 2007) y la relación entre el deterioro de ecosistemas y pérdida de biodiversidad para poder así formular estrategias de conservación más efectivas. Un mayor conocimiento de las bases de la resistencia de los ecosistemas frente a presiones es imprescindible para actuar directamente, reforzar esta capacidad de resistencia y para, cuando los impactos ya se han producido, catalizar y reforzar la capacidad de los ecosistemas y especies para recuperarse.

El impulso del conocimiento científico es imprescindible para formular nuevas estrategias de conservación, pero no es suficiente. El éxito de cualquier estrategia pasa por la reducción de las presiones derivadas de la actividad humana. Nuestra sociedad está actuando de forma seriamente irresponsable, erosionando y dilapidando la base de capital natural sobre la que se asienta nuestra calidad de vida presente y el futuro de nuestra especie. El conocimiento científico ha de traspasar el ámbito de la comunidad científica para informar a la sociedad y contribuir a formar ciudadanos mejor informados y más responsables. Hemos de traspasar las fronteras del conocimiento y las que lo separan de nuestra sociedad. En el éxito en este empeño nos jugamos en buena medida nuestro futuro.

Bibliografía

- Andersen, T., J. Carstensen, E. Hernández-García, y C. M. Duarte. «Ecological Thresholds and Regime Shifts: Approaches to Identification. Trends In Ecology and the Environment». 2008 (en prensa).
- Bindoff, N. L., J. Willebrand, V. Artale, A. Cazenave, J. Gregory, S. Gulev, K. Hanawa, C. Le Quéré, S. Levitus, Y. Nojiri, C. K. Shum, L. D. Talley y A. Unnikrishnan. «Observations: Oceanic Climate Change and Sea Level. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change». En Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor y H. L. Miller (eds.). Cambridge: Cambridge University Press, 2007.
- Bouchet, P. «La magnitud de la biodiversidad marina». En C. M. Duarte (ed.), *La exploración de la biodiversidad marina: desafíos científicos y tecnológicos*. Madrid: Fundación BBVA, 2006: 32-64.
- Corliss, J. B., J. Dymond, L. I. Gordon, J. M. Edmond, R. P. Von Herzen, R. Ballard, K. Green, D. Williams, A. Bainbridge, K. Crane y T. H. Van Andel. «Submarine thermal springs on the Galapagos Rift». *Science* 203 (1979): 1.073-1.083.
- Costanza, R., R. D'arge, R. De Groot, S. Farber, M. Grasso, B. Hannon, K. Limburg et al. «The value of the world's ecosystem services and natural capital». *Nature* 387 (1988): 253-260.
- Crutzen, P. J. y E. F. Stoermer. «The "Anthropocene"». *Global Change Newsletter* 41 (2000): 12-13.
- Darwin, C. *Sobre el origen de las especies por la selección natural*. Londres: John Murray, 1859.
- Duarte, C. M. «The future of seagrass meadows». *Environmental Conservation* 29 (2002): 192-206.
- Eldredge, N. *Life in the Balance. Humanity and the Biodiversity Crisis*. Princeton: Princeton University Press, 1998.
- Fiala, G. y K. O. Stetter. «*Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100°C». *Archives of Microbiology* 145 (1986): 56-61.
- Fischlin, A., G. F. Midgley, J. T. Price, R. Leemans, B. Gopal, C. Turley, M. D. A. Rounsevell, O. P. Dube, J. Tarazona, A. A. Velichko. Ecosystems, their properties, goods, and services. Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, Eds. Cambridge: Cambridge University Press, 2007: 211-272.
- Harris, L. A., C. M. Duarte y S. W. Nixon. «Allometric laws and prediction in estuarine and coastal ecology». *Estuaries and Coasts* 29 (2006): 340-344.
- Holt, W. V., A. R. Pickard y R. S. Prather. 2004. «Wildlife conservation and reproductive cloning». *Reproduction* 127 (2004): 317-324.
- Jannasch, H. W. y M. J. Mottl. «Geomicrobiology of deep-sea hydrothermal vents». *Science* 229 (1985): 717-725.
- Jaume, D. y C. M. Duarte. «Aspectos generales de la biodiversidad en los ecosistemas marinos y terrestres». En C. M. Duarte (ed.), *La exploración de la biodiversidad marina: Desafíos científicos y tecnológicos*. Madrid: Fundación BBVA, 2006: 17-30.
- Karl, D. M., C. O. Wirsen y H. W. Jannasch. «Deep-sea primary production at the Galapagos hydrothermal vents». *Science* 207 (1980): 1.345-1.347.
- Lonsdale, P. «Clustering of suspension-feeding macrobenthos near abyssal hydrothermal vents at oceanic spreading centers». *Deep-Sea Research* 24 (1977): 857-863.
- May, R. M. «Thresholds and breakpoints in ecosystems with multiplicity of stable states». *Nature* 269 (1977): 471-477.
- Mayhew, P. J., G. B. Jenkins y T. G. Benton. «A long-term association between global temperature and biodiversity, origination and extinction in the fossil record». *Philosop. Transc. Of the Royal Society*, doi:10.1098/rspb. (2000): 1.302.
- Meehl, G. A., T. F. Stocker, W. D. Collins, P. Friedlingstein, A. T. Gaye, J. M. Gregory, A. Kitoh, R. Knutti, J. M. Murphy, A. Noda, S. C. B. Raper, I. G. Watterson, A. J. Weaver y Z. C. Zhao. «Global Climate Projections. En: Climate Change 2007: The Physical Science Basis». En Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor y H. L. Miller (eds.). Cambridge: Cambridge University Press, 2007.
- Millennium Ecosystem Assessment, 2005a. *Ecosystems and Human Well-Being: Wetlands and water Synthesis*. Washington D.C.: World Resources Institute.
- , 2005b. *Ecosystems & Human Well-Being*. Vol. 1. Island Press.
- Rezende, E. L., J. E. Lavabre, P. R. Guimarães, P. Jordano, J. Bascompte. «Non-random coextinctions in phylogenetically structured mutualistic Networks». *Nature* 448 (2007): 925-928.
- Scheffer, M. y S. R. Carpenter. «Catastrophic regime shifts in ecosystems: linking theory to observation». *Trends in Ecology and Evolution* 18 (2003): 648-656.
- Schwartz, M. W., C. A. Brigham, J. D. Hoeksema, K. G. Lyons, M. H. Mills y P. J. Van Mantgem. «Linking biodiversity to ecosystem function: implications for conservation ecology». *Oecologia* 122 (2000): 297-305.
- Strange, C. J. «Facing the brink without crossing it». *Bioscience* 57 (2007): 920-926.
- Thomas, C. D. et al. «Extinction risk from climate change». *Nature* 427 (2004): 145-148.
- Trenberth, K. E., P. D. Jones, P. Ambenje, R. Bojariu, D. Easterling, A. Klein Tank, D. Parker, F. Rahimzadeh, J. A. Renwick, M. Rusticucci, B. Soden y P. Zhai. «Observations: Surface and Atmospheric Climate Change. En: Climate Change 2007: The Physical Science Basis». En Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor y H. L. Miller (eds.). Cambridge: Cambridge University Press, 2007.
- Vaquer-Sunyer, R., C. M. Duarte. «Thresholds of hypoxia for marine biodiversity». *Proceedings of the National Academy of Sciences* 105 (2008): 15.452-15.457.
- Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Ruchs, J. A. Eisen, D. Wu et al. «Environmental genome shotgun sequencing of the Sargasso Sea». *Science* 304 (2004): 66-74.

movilidad geográfica en un mundo condicionado por el clima

JOHN B. HEYWOOD

El contexto y el desafío ¹

La movilidad geográfica siempre ha sido un valor apreciado. Durante gran parte de la historia de la humanidad ha significado desplazar individuos o mercancías a la velocidad a la que una persona es capaz de caminar, un caballo avanzar, un buey tirar de un carro o un barco surcar el agua impulsado por remos o velas. Hasta el siglo XIX, en que aprendimos a explotar la energía de los hidrocarburos, no fuimos capaces de desplazar ni individuos ni mercancías a mucha más velocidad. La invención del vehículo a motor propulsado por petróleo a finales del siglo XIX y del avión a principios del XX abrió el camino a nuevas maneras de viajar e hizo posible una mayor velocidad en los desplazamientos. Las carreteras ofrecían una amplitud de itinerarios de la que el ferrocarril carecía y los aviones tan sólo precisaban vías aéreas para viajar de un lugar a otro.

Como resultado de estas innovaciones el siglo XX se convirtió en la «edad dorada» de la movilidad geográfica. El volumen de desplazamientos de personas y mercancías creció a un ritmo sin precedentes. Terminado el siglo, individuos cuyas vidas habían transcurrido hasta dentro de un radio de cien kilómetros de distancia de su lugar de nacimiento viajaban ahora hasta continentes remotos por negocios o por placer como si tal cosa. Las materias primas, los productos manufacturados y los alimentos de un extremo del mundo eran accesibles para los habitan-

tes del otro extremo. Las distintas regiones geográficas no participaron de forma igualitaria en esta expansión. Llegado el siglo XXI, para el ciudadano medio de un país rico las largas distancias se han convertido prácticamente en algo irrelevante. Los habitantes de los países pobres, sin embargo, continúan en gran medida desplazándose como lo hicieron sus antepasados.

Todo el mundo está a favor de la movilidad geográfica, tanto por el hecho en sí como porque les permite, de una parte, superar las distancias que separan sus hogares de los lugares donde trabajan, compran, van a la escuela o hacen negocios, visitan amigos o familiares, y de otra, conocer sitios nuevos. A los negocios les conviene la movilidad porque les ayuda a superar las distancias que los separan de sus fuentes de materias primas, de sus proveedores y sus mercados, y a evitar el impacto negativo de la congestión. Sin embargo, crece la preocupación por el hecho de que los sistemas de movilidad dependen hoy por hoy de una única fuente de energía: el petróleo. Y la tensión existente entre el deseo humano de movilidad y su preocupación por el impacto negativo asociado a la misma plantea cuestiones fundamentales relativas a su futuro.

Durante la segunda mitad del siglo XX las consecuencias negativas del aumento de la movilidad se hicieron evidentes a escala regional e incluso global. La contaminación generada por los motores de combustión inter-

¹
Fuente: WBCSD (Consejo Empresarial Mundial para el Desarrollo Sostenible, 2004).

na de cientos de vehículos causa la degradación de la calidad del aire en más y más ciudades. La prospección, extracción, transporte y refinamiento de petróleo para su empleo como combustible en medios de transporte daña el medioambiente en un grado cada vez mayor. El ruido que producen en la tierra y en el aire los vehículos que transportan personas y mercancías interrumpe la tranquilidad de millones de personas. Y las emisiones de dióxido de carbono procedente de la combustión de hidrocarburos, gran parte de la cual está relacionada con el transporte, están afectando al clima de nuestro planeta.

Nos vemos por tanto obligados a preguntarnos si el auge de la movilidad geográfica que ha definido los cincuenta últimos años es *sostenible*. El Consejo Empresarial Mundial para el Desarrollo Sostenible o WBCSD define *movilidad sostenible* como «la capacidad de satisfacer las necesidades de la población de desplazarse con libertad, comunicarse, comerciar y establecer relaciones sin sacrificar por ello otros valores esenciales humanos y ecológicos hoy o en el futuro» (WBCSD 2001). De acuerdo con esta definición, los usos actuales de medios de transporte no son sostenibles.

Dicho sencillamente, somos demasiados, consumimos demasiados recursos y lo hacemos de maneras que están dañando irreversiblemente nuestro planeta. El consumo es excesivo para la salud de la Tierra y, además, crece día a día debido al aumento demográfico, a las progresivas afluencia, urbanización y suburbanización y al continuo aumento de las expectativas de los consumidores. Y sin embargo es un hecho casi universalmente reconocido que la movilidad es uno de los elementos más importantes de un nivel de vida deseable.

La mayoría de los ciudadanos de los países ricos estamos satisfechos con nuestros sistemas de transporte, y gran parte del resto de la población mundial aspira a tener lo que nosotros tenemos. Pero cada vez somos más conscientes del precio de la movilidad, el cual incluye el desembolso económico que los usuarios hacen a los proveedores de la misma para que éstos, a su vez, puedan costearla. Pero la cosa va más lejos. El aumento de la movilidad ha venido acompañado de congestión circulatoria, riesgo de muerte y accidente, alteración de ecosistemas y comunidades, contaminación terrestre y del agua y emisión de gases de efecto invernadero.

El Consejo Empresarial Mundial para el Desarrollo Sostenible puso en marcha un gran proyecto llamado «Movilidad en 2030: Los desafíos a la sostenibilidad» (WBCSD 2004), que identificaba siete grandes objetivos para hacer sostenible el transporte:

1. Asegurarse de que las emisiones de agentes contaminantes convencionales relacionados con el transporte no constituyen un riesgo significativo para la salud pública en ningún lugar del mundo.
2. Limitar la emisión de gases de efecto invernadero (GEI) relacionados con el transporte a niveles sostenibles.

3. Reducir de forma significativa el número global de muertes y lesiones graves por accidentes de tráfico tanto en los países desarrollados como en los países en vías de desarrollo.
4. Reducir la contaminación acústica relacionada con los transportes.
5. Mitigar la congestión circulatoria.
6. Paliar las diferencias de acceso a la movilidad que impiden a los habitantes de los países más pobres y a los miembros de grupos sociales económicamente desfavorecidos mejorar sus condiciones de vida y las de sus familias.
7. Preservar y aumentar el acceso a la movilidad para la población en general, tanto en países desarrollados como en vías de desarrollo.

Es una lista de objetivos extremadamente ambiciosa. Nuestro desafío es obtener progresos en algunos puntos concretos sin perder de vista el contexto general. Para hacer frente a este reto resulta de ayuda plantearnos la pregunta: ¿cuáles son los elementos verdaderamente *no sostenibles* de los medios de transporte actuales? Hace una década participé en la elaboración de un estudio de la US National Academies sobre este tema (NRC 1997), que concluía que los factores de insostenibilidad eran dos: el primero era el riesgo de cambio climático causado por las emisiones de CO₂, a las cuales el transporte contribuye de manera significativa. El otro era la degradación de los ecosistemas y la reducción de la biodiversidad resultantes de las ya mencionadas emisiones y del impacto de las infraestructuras. Se trata de dos asuntos fundamentales porque la movilidad creciente está agravándolos de forma inevitable. Ambos están relacionados con las exigencias de energía del transporte. En este ensayo explicaré cómo podemos reducir el consumo de energía de los medios de transporte y la emisión de gases de efecto invernadero y, en consecuencia, su impacto en el cambio climático. Para ello no basta centrarse en el segundo objetivo del WBCSD de reducir las emisiones de GEI. Al mismo tiempo debemos tratar de que se cumplan los objetivos seis y siete, porque el aumento de la movilidad es esencial para asegurar el crecimiento económico en todas las regiones del mundo. Y también debe avanzarse en el resto de los objetivos si la mejora en los niveles de movilidad ha de seguir siendo un factor decisivo en el progreso económico y social.

Tamaño, crecimiento y complejidad

Nuestros sistemas de transporte en el mundo desarrollado desplazan a personas en automóvil, autobús, tren y avión. En los países en desarrollo también se usan la bicicleta y otros vehículos de dos y tres ruedas. El transporte de mercancías se hace principalmente por carretera y ferrocarril a partes iguales, aunque el transporte aéreo aumenta con rapidez. Los camiones pesados son los más empleados en el transporte por carretera. Los sistemas de transporte pue-

den ser de escala urbana, nacional o regional. Las figs. 1 y 2 muestran el estado actual y la proyección a futuro de las estadísticas básicas por región y modalidad de transporte tanto de personas como de mercancías. El crecimiento continuado que se produce cada año a un ritmo de un pequeño tanto por ciento es evidente. En la actualidad las regiones desarrolladas y en vías de desarrollo del mundo son comparables en escala, pero los índices de crecimiento en los países en desarrollo son más altos. Estas proyecciones (WBCSD 2004) están basadas en gran medida en el crecimiento demográfico y en la renta per cápita. Para 2050 se espera que estas estimaciones se hayan duplicado.

Las cifras indican la gran magnitud que ha adquirido el transporte: el número de vehículos en uso, los kilómetros que recorren y el peso de las mercancías que transportan. Con una población mundial de 6,8 millones y 800 millones de vehículos, la distancia media recorrida es de 5.000 kilómetros por persona y año (en Estados Unidos la media es de 20.000 kilómetros por persona y año y en África de 3.000). En el presente son los países desarrollados los que dominan el uso de vehículos, pero pronto los alcanzarán grandes áreas del mundo en vías de desarrollo. El transporte de mercancías es de 8 toneladas por persona y día. El consumo de combustible en los transportes se acerca a los 3.500 litros por persona y año, del cual casi la mitad es gasolina, un tercio es diesel y una sexta parte combusti-

ble para aviones. Todo ello cifra la producción total mundial de petróleo en alrededor de 92 barriles al día (cada barril contiene 159 litros). Estas tasas de consumo son tan grandes que resultan inabarcables. Pero además crecen a una tasa anual de aproximadamente el 2%.

¿Por qué preocuparnos del futuro? y, sobre todo, ¿cómo está afectando nuestro consumo energético a nuestro entorno? La razón es el tamaño de estos sistemas, su crecimiento aparentemente inexorable y el daño que hacen al medioambiente. Funcionan con combustibles derivados del petróleo (gasolina, diesel y combustible de aviones) a unas cantidades inimaginables. Cuando estos combustibles se queman en el interior de los motores el carbón presente en ellos se oxida y se convierte en el dióxido de carbono causante del efecto invernadero. La cantidad de dióxido de carbono que penetra en la atmósfera como resultado de la quema de estos combustibles es inmenso. Los transportes son responsables del 25% de las emisiones mundiales de gases de efecto invernadero. Conforme los países en vías de desarrollo se motorizan, cosa que está ocurriendo a gran velocidad, la demanda global de combustible supondrá un problema. También aumentará de forma significativa la concentración de gases de efecto invernadero en la atmósfera. Los vehículos ligeros de Estados Unidos (automóviles, furgonetas, todoterrenos y camionetas), cuya cantidad asciende a unos 250 millones,

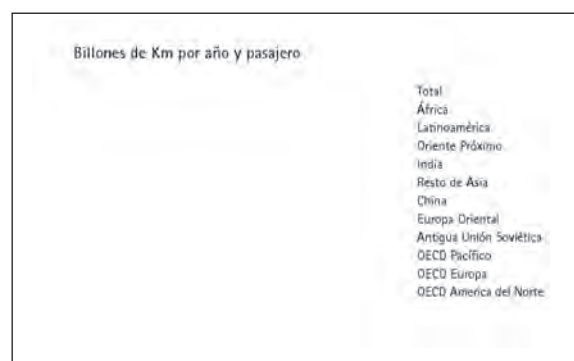


Figura 1(a)

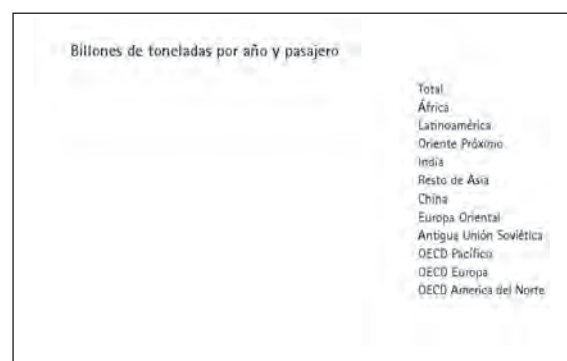


Figura 2(a)

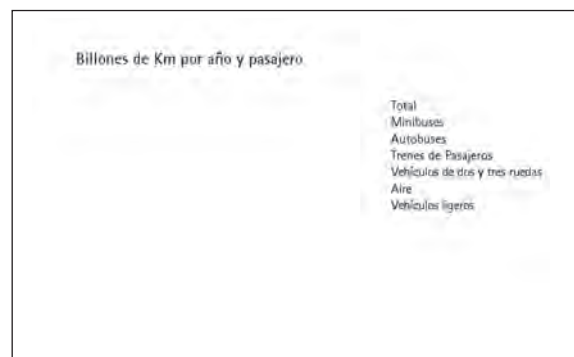


Figura 1(b)

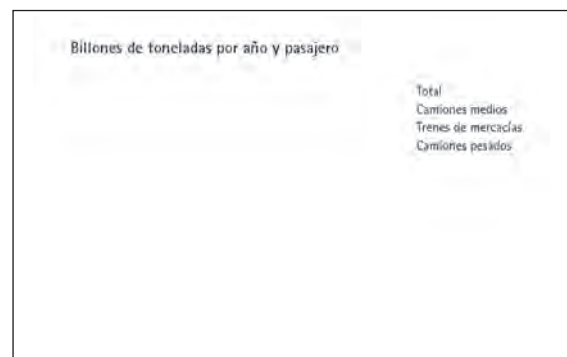


Figura 2(b)

Figura 1. (a) Movilidad por individuo y región; (b) movilidad por individuo y modalidad de transporte; proyección hasta 2050 (WBCSD 2004).

Figura 2. (a) Transporte de mercancías por región; (b) Transporte de mercancías por modalidad; proyección hasta 2050 (WBCSD 2004).

consumen en la actualidad 600.000 millones de litros de gasolina por año. Si otros países consumieran gasolina al mismo ritmo, el consumo mundial crecería en casi un 10%.

Varios países han empleado estándares de emisión de CO₂ para reducir el consumo de energía en los transportes y la emisión de GEI. Puesto que los hidrocarburos derivados del petróleo tienen un 87% de carbón en su composición, su consumo y las emisiones de CO₂ resultantes están directamente relacionadas: la combustión de 1 kilo de combustible libera 3,2 kilos de CO₂. Se espera que la reducción de consumo establecida por las nuevas regulaciones sea del 30% en 2020; dentro de más de 10 años. En 25 años (es decir, para 2035), se cree posible llegar a un factor de reducción 2. Mirando más hacia el futuro, a 2050, las estimaciones indican que hará falta al menos una reducción del 70% de los niveles de emisión de GEI para mantener los niveles de CO₂ de la atmósfera por debajo de las 550 ppm (partes por millón), una concentración considerada por muchos como la mejor a la que podemos aspirar en nuestros intentos por detener el calentamiento global. Todos estos objetivos, a corto, medio y largo plazo, presentan numerosos desafíos, porque los cambios necesarios para que se cumplan (ya sea el desarrollo de mejores tecnologías o la implementación de medidas de conservación) a *tamaño escala* requieren grandes cantidades de esfuerzo, tiempo y dinero.

Nuestras opciones de cambio

Si miramos al futuro, ¿qué oportunidades tenemos para hacer los medios de transporte más sostenibles a un coste aceptable? Existen varias opciones que supondrían una diferencia fundamental. Podríamos mejorar o cambiar la tecnología de los vehículos para hacerla más eficiente; podríamos cambiar la manera en que usamos nuestros vehículos para consumir menos combustible; podríamos reducir el tamaño y el peso de nuestros vehículos; podríamos usar otros tipos de combustible con menores huellas de carbono. Es muy probable que tengamos que adoptar

todas estas medidas si queremos llegar a la reducción drástica de la energía consumida en transportes y a la emisión de gases de efecto invernadero que se considera necesaria.

Al examinar las alternativas debemos tener presente los siguientes aspectos de nuestro actual sistema de transportes. En primer lugar, se adapta bien a su contexto primario, el de proporcionar movilidad en el mundo desarrollado. El paso del tiempo le ha permitido evolucionar de forma que mantiene un equilibrio económico entre sus costes y las necesidades de sus usuarios. En segundo lugar, este sistema vasto y optimizado depende por completo de una sola fuente de energía: el petróleo. Y ha desarrollado tecnologías —motores de combustión interna en la tierra y motores de inyección (turbina de gas) para el aire— de manera que el funcionamiento de los vehículos se adapta perfectamente a las propiedades de este combustible. Por último, los vehículos son duraderos, por lo que la reducción de su impacto medioambiental se prolonga durante mucho tiempo. Limitar y después reducir el impacto local y global de la energía empleada en transportes llevará décadas.

Fijémonos en la eficacia con la que empleamos energía en nuestros vehículos. Los índices de eficacia pueden llevar a malentendidos: lo que cuenta en realidad es el combustible que se consume cuando se está conduciendo. La figura 3 muestra el flujo de energía en un típico turismo de tamaño medio que circula en ciudad. Sólo alrededor del 16% de la energía generada por el combustible impulsa de hecho las ruedas: sus otras funciones son contrarrestar la resistencia aerodinámica, la del rodamiento de los neumáticos y acelerar el vehículo. El consumo de combustibles puede limitarse reduciendo las pérdidas en el sistema de propulsión (Kasseris y Heywood 2007). El motor a gasolina actual tiene una eficiencia de alrededor de un 20% en la circulación urbana, aunque puede llegar al 35% en condiciones óptimas. Pero muchos desplazamientos cortos con la transmisión y el motor fríos, sumados al impacto del clima invernal y a una conducción agresiva empeo-

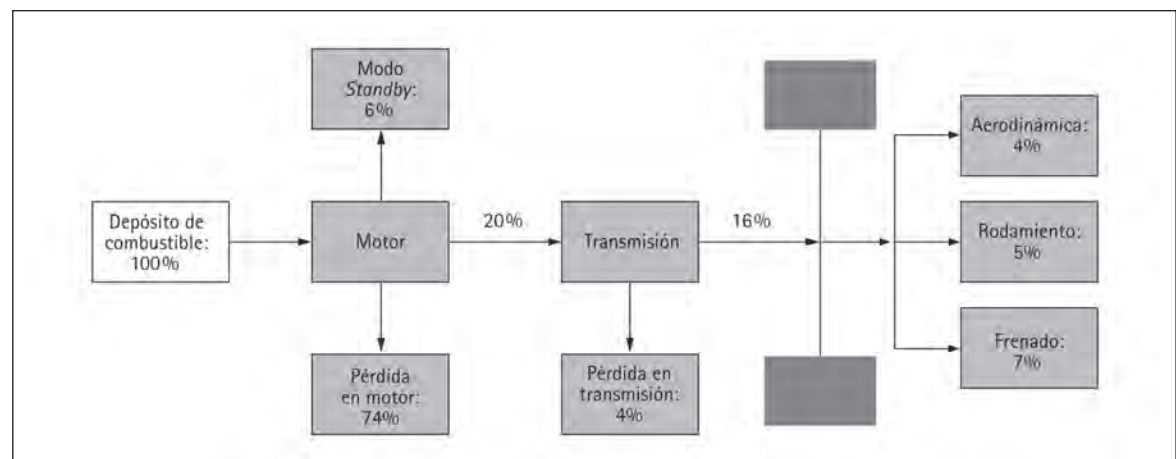


Figura 3. La energía fluye desde el depósito del vehículo a las ruedas en un coche de pasajeros típico en un área urbana (Bandivadekar et al. 2008).

ran de manera significativa el consumo de combustible, como también lo hace el tiempo que se pasa con el coche detenido pero con el motor en marcha. Estos aspectos de la conducción real reducen de hecho el índice medio de eficiencia, de manera que tan sólo un 10% aproximadamente de la energía química almacenada en el combustible sirve de hecho para mover las ruedas. Amory Lovins, un gran defensor de los vehículos más ligeros y eficientes, lo explica de esta manera: con un vehículo de un 10% de eficiencia y con un conductor, un pasajero y equipaje —lo que supondría una carga útil aproximada de 140 kilos, un 10% del peso del vehículo— sólo un 1% del combustible del vehículo se emplea para mover la carga útil (Lovins et al. 2005). No hay duda de que esto puede mejorarse.

Cuando hacemos el cálculo de los índices de emisión de gases de efecto invernadero generada por el uso de vehículos debemos incluir la energía necesaria para producir el combustible a partir del petróleo en crudo y después distribuirlo, conducir el vehículo durante su vida media (de 150.000–240.000 kilómetros), así como la fabricación, el mantenimiento y la posterior destrucción del vehículo. Estas fases se conocen con los nombres de *well-to-tank* (eficiencia desde el pozo al depósito), que equivale a alrededor del 15% del consumo total de energía y emisiones de GEI, *tank-to-wheel* (del depósito a las ruedas, 75%) y *cradle-to-grave* (literalmente de la cuna a la tumba, aunque sería más apropiado decir de la fábrica al desguace), que supone un 10%. Vemos que la energía requerida para producir el combustible deviene especialmente importante cuando tomamos en consideración combustibles no derivados del petróleo, como los biocombustibles o el hidrógeno, así como los nuevos sistemas de propulsión de vehículos, pues lo que importa aquí es lo que realmente se usa y lo que se emite.

Enseguida explicaré que mejorando la tecnología existente de los vehículos ligeros se puede conseguir mucho. Si se invierte más dinero en mejorar la eficiencia de los trenes de transmisión existentes, se reduce el peso de los vehículos y la resistencia aerodinámica y si se mejoran los neumáticos podremos reducir el consumo de combustible a alrededor de una tercera parte del actual durante los próximos 20 años, lo que supondría una mejora anual de alrededor del 3%. Esta reducción del consumo de combustible costaría unos 2.000 dólares por vehículo. Teniendo en cuenta lo que costarán los carburantes en el futuro, esta cantidad no incrementaría el coste general de poseer un vehículo. De hecho, el incremento de los costes que supone mantener un vehículo se ha producido de manera continuada durante los últimos 25 años, pero lo que sucede es que hemos invertido en coches más grandes, pesados y veloces y en camiones más ligeros, en lugar de hacerlo en eficiencia en el consumo de carburantes. Aunque resulta especialmente evidente en Estados Unidos, este cambio gradual a coches más grandes y potentes se da y sigue dándose en todas partes.

¿Cuáles son nuestras opciones en cuanto a motores o sistemas de propulsión? Podemos seguir empleando motores de ignición que funcionen con gasolina o con biocombustibles como el etanol. Los vehículos eléctricos híbridos o HEV (*hybrid electrical vehicles*), con motor de combustión interna y motor y batería eléctricos, son otra opción cada vez más viable. En la próxima década más o menos, los vehículos eléctricos híbridos recargables en la red podrían convertirse en una alternativa seria, que emplearía cualquiera de los carburantes líquidos existentes junto con electricidad para recargar sus baterías. Más a largo plazo, es posible predecir para 2030 el uso generalizado de vehículos de hidrógeno y de vehículos de batería eléctrica, pero ello podrá suponer un aumento de los costes. Además es probable que para entonces sean ya una realidad la reducción del peso de los vehículos y de la resistencia a los neumáticos y al aire, con lo que se mejorarán los sistemas de propulsión. Hay que señalar que la incorporación de un motor y una transmisión más pequeños, ligeros y eficientes a vehículos ligeros supondría combinar todos estos avances de maneras especialmente ventajosas.

Los motores de gasolina e ignición estándares siguen mejorando su potencia por unidad de volumen desplazado así como su eficiencia operativa media en aproximadamente un 2% por año. Estas mejoras proceden de la reducción de la fricción, del control variable de las válvulas del motor, de la desactivación de los cilindros cuando el motor no está muy cargado, de la inyección directa de combustible al cilindro del motor, del aumento de la tasa de compresión del motor y del uso de avanzados sensores que controlan el funcionamiento del motor. Se está aumentando el rendimiento del motor de gasolina incrementando la presión de aire mediante un turbocargador. Ello permite reducir el tamaño del tren de transmisión y que éste sea más eficiente. Los motores diesel también pueden mejorar su eficiencia, aunque no tanto como los de gasolina. Los que se fabriquen en el futuro deberán incorporar tecnología (purgadores y catalizadores) para reducir la emisión de gases en sus tubos de escape con objeto de controlar las partículas contaminantes y el óxido de nitrógeno (NO_x). Ello supondrá costes más elevados y penalización en el precio del combustible. Por lo tanto, los futuros motores de gasolina o diesel con turbocargadores serán más parecidos en cuanto a funcionamiento, potencia por volumen desplazado y eficiencia operativa media. Y, lo que es más importante, el futuro motor de gasolina será significativamente más económico —alrededor de la mitad de precio— que el diesel.

Los HEV que se fabrican y se venden hoy día suponen un porcentaje muy pequeño dentro del mercado. Los híbridos actuales están dotados de una batería, un motor eléctrico, un generador, controles de potencia eléctrica y un avanzado sistema de transmisión. Casi todas las configuraciones actuales emplean una combinación híbrida similar, en la que la transmisión tiene capacidad para des-

conectar tanto el motor de combustión como el eléctrico de las ruedas, y un sistema de gestión inteligente que apaga el motor cuando la carga es ligera o no hay carga, y que recupera hasta el 90% de la energía de frenado convirtiéndola en energía eléctrica. Estos «híbridos recargables» reducen de forma significativa el consumo de combustible, y la magnitud del ahorro dependerá del tipo de conducción (por ejemplo, si es conducción urbana a baja velocidad o de alta velocidad por autopista) además de otros detalles clave. En el futuro, cuando esté perfeccionada la tecnología híbrida, la reducción del consumo de combustible en motores de gasolina podrá alcanzar el 40 o el 50%. La «conducción eléctrica» reforzada por un motor de combustión interna ya constituye una atractiva opción de futuro. Es probable, sin embargo, que resulte entre 2.000 y 3.000 dólares más cara que el tren de transmisión convencional mejorado (Bandivadekar et al. 2008).

El vehículo híbrido conectable a la red (o PHEV, por sus siglas en inglés) funciona, como su nombre indica, con gasolina y electricidad y lleva una batería más grande que puede recargarse desde un enchufe o toma de corriente. Emplea un sistema de batería avanzado (por ejemplo, con tecnología de litio-iones) en una configuración similar a la del híbrido convencional. El PHEV supera el umbral de la batería de estado de carga o SOC y opera con un dispositivo en «modo de agotamiento» (CD, *charge-depleting*) que emplea la energía eléctrica de la batería para proveer de potencia al vehículo. Cuando llega el umbral mínimo de SOC, el vehículo cambia al modo de agotamiento, que equivale al del HEV convencional. Funciona con energía eléctrica y de combustión. Conviene señalar que la electricidad que usa el vehículo consume cerca de tres veces más que la energía primaria producida con carburantes. La tecnología de vehículos híbridos conectables a la red está aún en fase de desarrollo y resulta demasiado costosa para que pueda implantarse en el mercado. Además necesitará electricidad «verde» para que su contribución a la reducción de emisión de gases de efecto invernadero sea significativa o supere a la de los híbridos recargables convencionales.

Los vehículos de batería eléctrica o BEV obtienen toda la energía que requieren de la red eléctrica. Para su implantación será necesario un replanteamiento del tamaño, el coste y la gama de los vehículos, ya que la gama media de 650 kilómetros de autonomía con el depósito lleno no parece plausible en un vehículo completamente eléctrico desde el punto de vista de los costes y del peso; incluso pensar en 300 kilómetros parece poco viable. Por tanto los BEV no tienen de momento grandes posibilidades comerciales, aunque están concebidos fundamentalmente para uso urbano.

La célula de combustible para propulsión de vehículos incorpora un sistema de membrana de intercambio de protones (PEM) para alimentar un motor eléctrico, que hace funcionar el vehículo, por lo general con una confi-

guración en serie. Una célula de combustible opera como una batería, en el sentido de que transforma la energía química presente en el hidrógeno del carburante en electricidad. Su principal diferencia respecto a la batería es que la entrada de combustible (hidrógeno) y oxidante (aire) a los electrodos de la célula es continua. En una configuración híbrida de célula de combustible, la batería, que almacena energía eléctrica, mejora el rendimiento general del sistema y permite la regeneración del mismo durante la frenada. Este híbrido emplea la misma batería de alta potencia de litio-iones, que ahora empieza a incorporarse en los vehículos híbridos convencionales. Los vehículos de célula de combustible deben todavía superar numerosos obstáculos tecnológicos y reducir significativamente sus costes antes de que puedan comercializarse a gran escala. Concretamente, el rendimiento y la durabilidad de la célula de combustible están limitados por las propiedades de los materiales de membrana de electrolito actuales, por los requerimientos del catalizador y por lo complicado de los sistemas de gestión necesarios para su funcionamiento. Además de ello hace falta desarrollar sistemas de almacenamiento del hidrógeno dentro de los propios vehículos para poder ampliar la gama de los mismos, lo que por el momento supone un problema de costes y peso. Por supuesto, producir y distribuir hidrógeno, es decir, crear una infraestructura de este combustible, supone también un gran desafío.

Los sistemas de transmisión de los vehículos son también muy importantes. Los automáticos son muy populares en Estados Unidos, sobre todo debido a su facilidad de uso, y esta popularidad se está extendiendo a otros países. Es muy probable que su eficiencia mejore en un periodo de corto a medio plazo al incrementar en número de marchas y reducir así la pérdida de energía por el mecanismo de cojinetes y gracias a elementos de sellado y sistema hidráulico. Aunque las marchas de cuatro cambios siguen dominando el mercado estadounidense, las de cinco se están volviendo habituales. Las de seis velocidades y las manuales automatizadas existen en algunos modelos de coches y es probable que su uso se estandarice en la próxima década. Los vehículos de lujo han empezado además a incorporar cajas de cambio de siete y ocho marchas, que podrían ser habituales a medio plazo. Por tanto para el futuro se predice una eficiencia en el sistema de cambio de marchas de entre el 90 y el 95%. Se trata de una mejora significativa en comparación con la anterior generación de sistemas de transmisión.

La reducción del peso de los vehículos es otra forma obvia de mejorar el consumo de combustible. La figura 4 ilustra las necesidades de combustible de un vehículo ligero de tamaño medio en Estados Unidos. La regla general es que una reducción del 10% en el peso del vehículo puede reducir el consumo de carburante en un 5-7% si se acompaña de una reducción proporcional del tamaño del motor, de materiales de fabricación más ligeros y de un rediseño

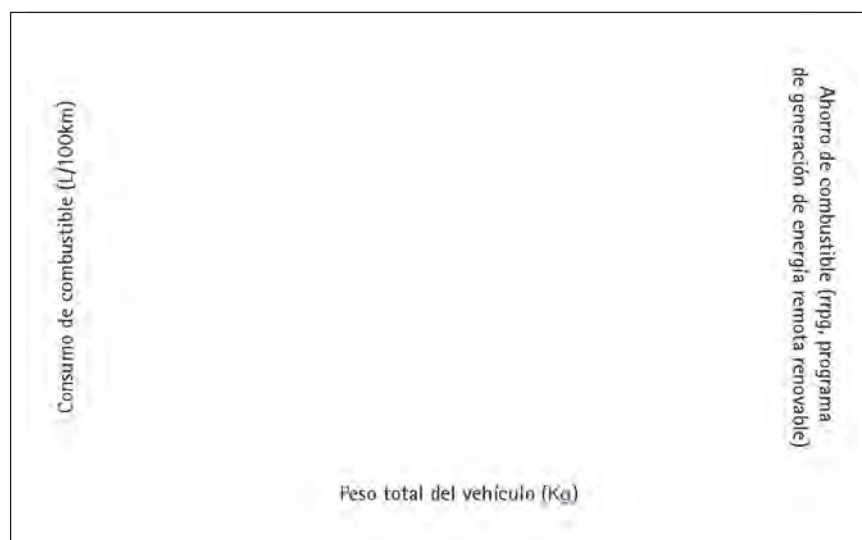


Figura 4. Relación entre consumo de combustible y peso en vehículos ligeros basado en los modelos de vehículos estadounidenses en 2005 (Cheah et al. 2008).

adecuado del coche. Eliminar una plaza de un automóvil de pasajeros puede reducir su peso en un 10%, aproximadamente. Sin embargo, el tamaño de los coches es una de las prestaciones que los consumidores más valoran.

La reducción de la resistencia al rodado de los neumáticos es otra manera de ahorrar combustible. No olvidemos que las nuevas tecnologías de fabricación de neumáticos pueden introducirse primero en el mercado de recambio de ruedas y por tanto obtener resultados más rápidamente que si se implementan sólo en la fase de fabricación del coche. La supervisión de la presión de los neumáticos y que éstos tengan el aire adecuado también son factores que contribuyen al consumo eficiente.

La reducción de emisión de GEI en vehículos de gasolina y diesel con motor de combustión interna y en HEV es proporcional a la reducción en consumo de petróleo. Esta reducción podría ir más allá, si el contenido en carbón de los combustibles pudiera rebajarse empleando biocombustibles con niveles bajos de carbón. Para los PHEV, los BEV y los FCV, las emisiones «del pozo al depósito» producidas durante la generación y el suministro de electricidad e hidrógeno afectan en gran medida el potencial de reducción en la emisión de GEI. La mejora de la eficiencia en la producción de electricidad, así como las cada vez mayores contribuciones procedentes de fuentes de energía nucleares y renovables y de hidrocarburos con captura y almacenamiento de carbono podrían muy bien disminuir las mencionadas emisiones *well-to-tank* debidas a la generación de electricidad. De esta manera, los híbridos conectables a la red se convertirían en una alternativa atractiva a la hora de limitar las emisiones GEI y de petróleo.

Examinemos con mayor detalle nuestras opciones en cuanto a combustibles se refiere. Nuestros sistemas de transporte actuales —por tierra, mar y aire— han empleado de forma extensiva hidrocarburos derivados del petró-

leo. Ello se debe a que son combustibles líquidos, tienen una alta densidad de energía y encajan bien con las tecnologías actuales: motores de ignición, diesel y turbinas de gas. Un ejemplo de su atractivo para la industria de la automoción es que cada vez que repostamos hoy nuestros coches, la energía de los combustibles fluye por la manguera que sostenemos en la mano a una velocidad de 10 megavatios, lo que asegura otros 650 kilómetros de conducción y habiendo empleado sólo 5 minutos en repostar.

Puesto que los combustibles derivados del petróleo dominan el sector de transportes, han desarrollado unos sistemas de refinado y distribución a gran escala. En Estados Unidos se distribuye anualmente más de 1.365.000 de litros de derivados refinados del petróleo, lo que supone un tercio de la producción mundial. La capacidad de fuentes de combustible alternativas de ser compatibles e integrarse en estos sistemas de refinado y distribución es por tanto un aspecto clave de su viabilidad.

¿Cuáles son las alternativas posibles? El uso de gas natural en transporte varía desde menos del 1% en vehículos de la gran mayoría de los países hasta el 10% en un par de ellos, donde las políticas tributarias lo han convertido en una alternativa económica. El gas natural presenta características favorables a su empleo en el motor de combustión, pero se trata de un combustible gaseoso que debe comprimirse y después almacenarse en tanques de alta presión dentro del vehículo. Los inconvenientes de un combustible gaseoso (menor potencia específica del motor, gama de conducción limitada, tareas de compresión en el momento de repostar, necesidad de llevar a bordo tanques de almacenamiento con las limitaciones de espacio que ello supone, costes extra y emisión de metano) contrarrestan con mucho la principal ventaja de este combustible: su baja tasa de emisión de GEI.

Las arenas petrolíferas (por ejemplo, los depósitos alquitranosos de Canadá) y el petróleo pesado (más denso que el normal, se concentra en Venezuela) ya constituyen una fracción creciente (de alrededor del 5%) de los combustibles empleados en transportes. Con el tiempo, otras fuentes no petrolíferas de otros combustibles fósiles, tales como la conversión de gas natural a estado líquido, el aceite de esquisto bituminoso y el carbón, serán una realidad. Estas alternativas pueden ser combustibles de alta calidad y se espera que su volumen de producción crezca de forma continuada. Sin embargo, las emisiones de dióxido de carbono que se liberan durante su producción son más altas que las de la producción basada en derivados del petróleo, debido a la cantidad de energía requerida para obtenerlos y a su elevado contenido de carbono.

Los combustibles líquidos derivados de biomasa tienen potencial para contribuir de forma significativa al suministro de energía para el transporte. Fuentes de biomasa incluyen el maíz, las hierbas de la pradera, el pasto varilla, el miscanto, los desechos forestales y urbanos y otros cultivos específicamente orientados a la fabricación

de biocombustibles. Los productos manufacturados finales incluyen el metanol, el biodiesel y —es probable que en el futuro— los combustibles análogos a la gasolina y el diesel. Antes de su implantación es necesario resolver cuestiones cruciales como la disponibilidad de tierras de cultivo adecuadas, las emisiones de gases de efecto invernadero consecuencia del cambio de uso de la tierra de cultivo, las necesidades de agua y fertilizantes, la degradación de la tierra con el paso del tiempo, la contaminación de las aguas y las necesidades energéticas durante la producción. Existe un potencial importante en la contribución de los biocombustibles al transporte, pero su grado exacto aún debe someterse a extensos análisis. Dentro de 20 años, en Estados Unidos alrededor de un 30% de la energía empleada en transportes podría proceder de biocombustibles.

Los biocombustibles, la electricidad y el hidrógeno requieren distintos análisis del ciclo de producción y distribución de combustibles, que hoy constituye un factor clave. Los impactos de los biocombustibles varían: pueden ser desde, como en el caso del etano hecho de maíz, comparables en cuanto a emisión de gases GEI durante el ciclo completo de vehículos de gasolina, a inferior que el de éstos (el etanol fabricado a partir de caña de azúcar en Brasil) y a mucho menor cuando se trata de un combustible basado en la conversión en celulosa de biomasa. La carga que pueden suponer la necesidad de energía eléctrica y la emisión de GEI varían, puesto que dependen de cómo se genera la electricidad. Cuando se genera a partir de hidrocarburos la carga que supone la electricidad es sustancial, y el uso mucho más eficiente de la electricidad en el vehículo queda anulado por la ineficiencia procedente de la generación de electricidad en la planta eléctrica y en la distribución. Las cuestiones que importan aquí son: ¿cuáles son las fuentes viables de electricidad verde —es decir, de baja emisión de GEI— para el transporte con, digamos, híbridos conectables, y cuándo estará disponible dicha electricidad? El hidrógeno plantea inte-

rogantes similares: ¿cómo puede producirse y distribuirse con bajos niveles de emisión de GEI? Cualquier hidrógeno que se produzca a corto plazo habrá de proceder casi obligatoriamente del gas natural, cuyos índices actuales de consumo de energía y de emisión de GEI no difieren gran cosa de los resultantes del empleo de vehículos con combustibles derivados del petróleo.

Rendimiento de estas tecnologías de transporte

Hemos hecho proyecciones del rendimiento y los costes de todas estas tecnologías de propulsión para vehículos a 25 años. La figura 5 muestra las proyecciones para los principales vehículos con tren de transmisión (Bandivadekar et al. 2008). Se prevé un consumo de combustible más eficiente (con un rendimiento y un tamaño del vehículo constantes), pero también un aumento de los costes. La reducción media del tamaño del vehículo (del 20%) costará unos 700 dólares. En Europa y Asia el tamaño medio de los vehículos equivale a dos tercios del de Estados Unidos. Además, en Europa aproximadamente la mitad de los vehículos de pasajeros son diesel, de manera que la eficiencia media de consumo de combustible es ya mayor.

En general estos futuros vehículos más eficientes con sus diferentes trenes de transmisión costarán unos 2.000 dólares más que los de gasolina, entre 2.500 y 3.000 los vehículos de turbogasolina, entre 3.500 y 4.300 los diesel y entre 4.500 y 5.500 los híbridos, todos en comparación con sus actuales equivalentes convencionales, de motor de gasolina. Los híbridos conectables a la red y los vehículos de célula de combustible costarán probablemente entre 6.000 y 8.000 dólares más. En el momento presente el concepto de vehículo con sistema de batería y una capacidad significativa de almacenamiento interno no está aún listo para su implantación generalizada en el mercado.

En los vehículos de tecnología convencional, el ahorro de combustible que estos nuevos sistemas de propulsión hacen posible compensaría el aumento en los costes, tomando como referencia el precio actual de los carburantes. Pero a fecha de hoy este incentivo sólo ha resultado en un menor coste de las mejoras tecnológicas aplicadas a los vehículos y todavía no ha creado un mercado fuerte y en expansión para tecnologías nuevas y más eficientes tales como los vehículos híbridos, aunque los elevados precios del combustible y el hecho de que el diesel está gravado con menos impuestos que la gasolina, además del perfeccionamiento de la tecnología diesel, han hecho que esta opción suponga el 50% del nuevo mercado automovilístico de Europa.

Por lo tanto es importante completar el análisis ciclo de vida de un vehículo teniendo en cuenta el consumo de energía y las emisiones de GEI tanto durante su fabricación como durante su uso. El ciclo de producción del vehículo supone en la actualidad un 10% de la energía y de la emisión de GEI. Dentro de unos 25 años habrá ascendido al 15-20% debido al uso creciente de nuevos mate-



Figura 5. Consumo de combustible en los vehículos actuales y futuros con diferentes trenes de transmisión avanzados para 2006, 2020 y 2035 (Bandivadekar et al. 2008).

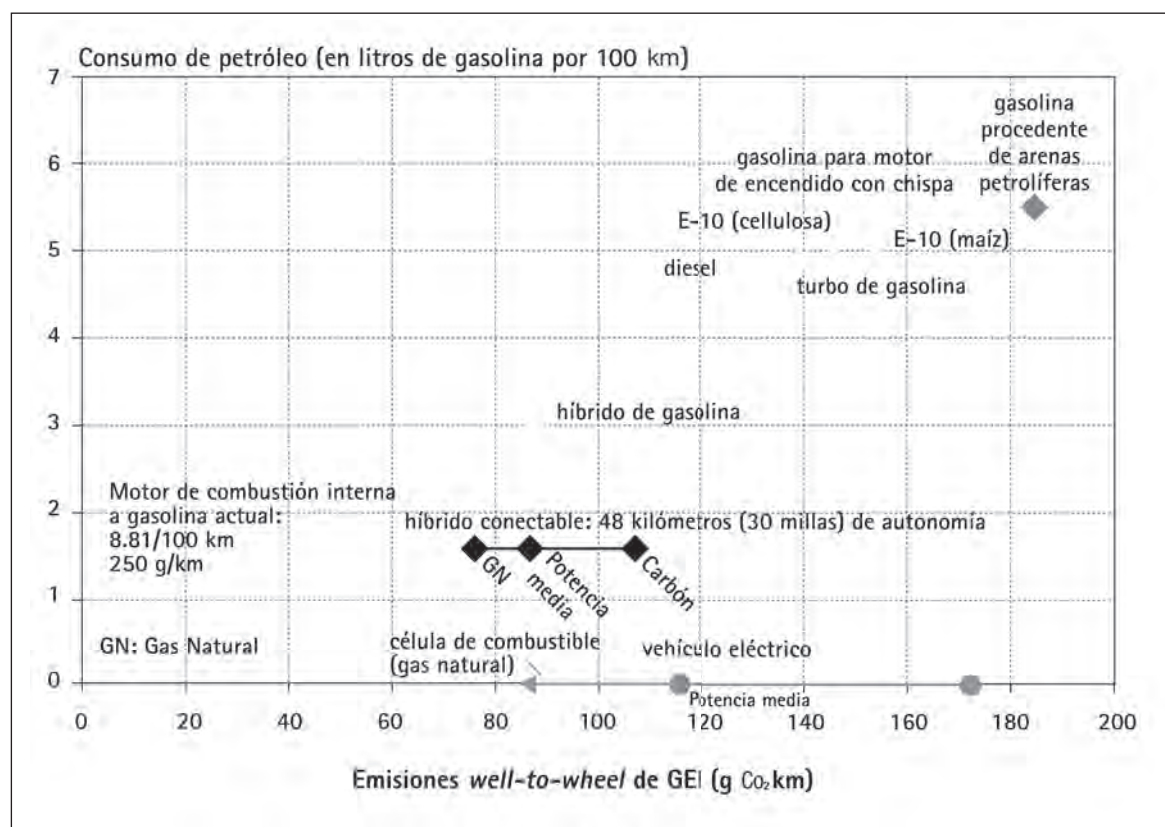


Figura 6. Consumo de petróleo y emisiones *well-to-wheel* de GEI de los futuros coches (2035) con sistemas de propulsión y fuentes de combustible diferentes (Bandivadekar et al. 2008).

riales más ligeros y más intensivos desde el punto de vista energético, así como a la reducción en el consumo de carburantes. El ciclo de producción y distribución de carburantes procedentes de hidrocarburos supone alrededor de un 20% del consumo de energía y emisión de GEI; los combustibles de hidrocarburos procedentes de fuentes petrolíferas no convencionales, como las arenas petrolíferas, supondrán aproximadamente el doble.

La figura 6 ilustra una comparación entre el consumo de petróleo como carburante y las emisiones *well-to-wheel* de GEI de estos distintos sistemas de propulsión del futuro, en un automóvil estadounidense de peso medio y con una proyección a 25 años. En lo que se refiere a consumo de combustible, los vehículos híbridos, eléctricos y de célula de combustible revelan una reducción significativa. Pero, claro está, estos vehículos también consumen electricidad de la red o hidrógeno procedente de una nueva infraestructura de producción y distribución. Estos requerimientos energéticos adicionales anulan la reducción de emisiones de GEI. Con sistemas de suministro eléctrico derivados del carbón y del gas natural (como los existentes en Estados Unidos), las emisiones de GEI adicionales están compensadas por las emisiones de la generación de electricidad. Con generación de energía eléctrica nuclear y renovable el panorama se vuelve más halagüeño. Las emisiones de un coche con batería eléc-

trica son peores en gran parte debido al peso extra que suponen las baterías. Como vemos aquí, las futuras emisiones de GEI por vehículo podrían con tiempo reducirse a una tercera parte de las que se dan actualmente. Las mejoras en los motores convencionales, en los sistemas de cambio de marchas y la reducción de la longitud y el peso de los vehículos, logrará disminuir el consumo de petróleo en un 30 o 40%. Podría obtenerse un 10% adicional mediante la utilización de cantidades plausibles de biocombustibles. La tecnología híbrida, por su parte, aporta una significativa reducción del 40%. Aunque los híbridos conectables y los de células de combustible con hidrógeno reducen significativamente e incluso eliminan el consumo de petróleo, en la fase de transición necesaria para su implantación su impacto en cuanto a emisión de gases de efecto invernadero no es mucho menor que el de los híbridos convencionales de gasolina.

Ventajas y comerciabilidad

Hasta el momento hemos comparado las características entre vehículos tomando como referencia un rendimiento constante y un tamaño fijo; es decir, cuando proyectamos estos vehículos a futuro sus características no varían. Los datos que poseemos de las dos últimas décadas demuestran que el rendimiento y el tamaño de los vehículos han aumentado a un ritmo constante, en espe-

cial el rendimiento. En Estados Unidos, mientras que los trenes de transmisión de los vehículos se han vuelto cada vez más eficientes en los últimos 20 años, el consumo de combustible en carretera se ha mantenido constante. En Europa, la mejora del rendimiento no ha sido tan grande y alrededor de la mitad de los avances en cuanto a eficiencia del motor se han traducido en reducción de consumo de combustible. Es esencial concentrar esfuerzos en reducir el consumo actual de combustible, y en el caso de Estados Unidos puede decirse que, desde principios de la década de 1980 estos esfuerzos han sido prácticamente inexistentes, mientras que aproximadamente la mitad de los beneficios potenciales del ahorro de combustible corresponden a Europa. Los compradores y los usuarios de vehículos han demostrado una clara preferencia por un mejor rendimiento y un mayor tamaño, causando que el mercado se centre en ofrecer estas prestaciones. Los fabricantes de automóviles compiten entre sí para ofrecer un rendimiento y un tamaño cada vez mayores. En Estados Unidos el énfasis en el alto rendimiento ha sido tal que, junto con aumentos en el tamaño de los coches, ha generado una pérdida en la reducción del consumo de combustible de cerca del 25%. En Europa la demanda de alto rendimiento de los coches no ha sido tan fuerte, y por eso se ha llegado al 50% de la reducción considerada plausible en el consumo de combustible.

Ya hemos dicho que el peso y el tamaño del vehículo pueden también contribuir de forma significativa a reducir el consumo de petróleo y la emisión de gases de efecto invernadero. Se trata de una oportunidad importante y una ventaja a sumar a las que supondrá la optimización de los trenes de transmisión de los vehículos. La reducción directa del peso mediante la utilización de materiales más ligeros y cambios en el diseño básico de los vehículos (que, por ejemplo, maximicen el volumen interior para obtener una longitud y una anchura determinadas) hacen posible nuevas reducciones de peso si se limita el tamaño de los componentes de fabricación. Un cambio en el tamaño de los vehículos también reduciría el

peso medio y puede lograrse en principio cambiando los volúmenes de producción de los modelos existentes en el mercado. Nuestros cálculos indican que podría lograrse un reducción del 20% del peso medio de los vehículos a lo largo de 25 años y a un coste de 700 dólares. El potencial máximo para reducción de peso en Estados Unidos es de cerca del 35%, pero esta reducción adicional del 20% sería significativa. Estamos hablando de reducciones sustanciales en el peso de los vehículos que requieren replantearse el diseño automovilístico. Una reducción del 20-35% por ciento del peso de los vehículos supondría una reducción de 12-20% del consumo de combustibles (Bandivadekar et al. 2008).

Penetración en el mercado: una cuestión fundamental

Las mejoras en los sistemas de propulsión y en la tecnología de los vehículos sólo tienen verdadera repercusión cuando se aplican en grandes cantidades. Estas tecnologías perfeccionadas deben por tanto poseer un gran atractivo comercial, su volumen de producción debe ser alto y mantenerse así durante 5-10 años para que tengan un impacto duradero en el mercado automovilístico. Una vez se conocen los índices de ahorro de combustible y reducción de emisiones GEI que prometen estos nuevos vehículos, entonces puede estudiarse cómo evolucionan dichos índices una vez los nuevos modelos entran en el mercado y los viejos van desapareciendo. Lo que es importante, aunque difícil de calcular, es la tasa de penetración de mercado o la evolución en los volúmenes de producción de estas nuevas tecnologías.

¿Qué es lo que dicta la introducción en el mercado de tecnologías de motor mejoradas y de combustibles alternativos? Incluso si la demanda de nuevos sistemas de propulsión está creciendo, las existencias podrían ser limitadas. Ello podría deberse a limitaciones en recursos de ingeniería y de capital, así como en las cadenas de suministros. El automóvil es un producto altamente complejo y las expectativas del consumidor cada vez son mayores. El desarrollo

Período de implementación	Tecnología de vehículos				
	Gasolina Inyección directa con turbocargador	Diesel de alta velocidad con catalizador para reducción de material particulado y óxido de nitrógeno (NOx)	Híbrido gasolina/ batería eléctrica	Híbrido recargable gasolina/batería eléctrica	Híbrido de célula de combustible con almacenamiento de hidrógeno a bordo
Vehículo competitivo en el mercado	~ 2-3 años	~ 3 años	~ 3 años	~ 8-10 años	~ 12-15 años
Penetración en la producción de nuevos vehículos	~ 10 años	~ 15 años	~ 15 años	~ 15 años	~ 20-25 años
Penetración en el mercado en general	~ 10 años	~ 10-15 años	~ 10-15 años	~ 15 años	~ 20 años
Total de tiempo requerido	~ 20 años	~ 25 años	25-30 años	~ 30-35 años	~ 50 años

Cuadro 1. Plazos estimados para el impacto de las nuevas tecnologías (adaptado de Schafer et al. 2006).

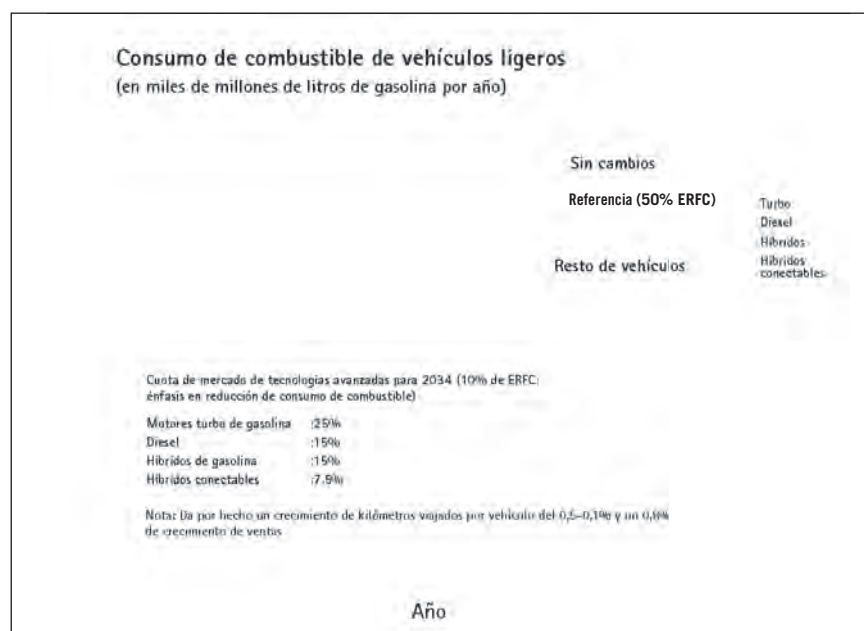


Figura 7. Panorama hipotético que examina el impacto en el consumo de combustible de la flota de vehículos de Estados Unidos de la presencia creciente de vehículos de tecnología avanzada hasta 2035. La mitad de los avances en cuanto a eficiencia (50% ERFC) están encaminados a reducir el consumo de combustible de los vehículos; la otra mitad, a contrarrestar el aumento en tamaño y en rendimiento. La reducción del peso de los vehículos es el 10%. El impacto en el tamaño de la flota y el kilometraje por vehículo no está reflejado (Bandivadekar et al. 2008).

y diseño de nuevos trenes de transmisión y otros subsistemas, incluida una nueva arquitectura del automóvil, es una tarea laboriosa y puede que transcurran 15 años antes de que esté al alcance de todos los segmentos del mercado.

La fabricación de automóviles es un negocio que requiere grandes cantidades de capital y mano de obra, y quienes lo emprenden a menudo asumen importantes riesgos. Por lo general, un fabricante de automóviles tarda de dos o tres años en montar una nueva planta industrial. Por tanto, reconvertir el 10% de la capacidad de producción doméstica de Estados Unidos (1,5 millones de vehículos al año) de vehículos híbridos supondría una inversión de aproximadamente 220.000 millones de dólares, el equivalente al 10% del gasto anual del sector de fabricación de automóviles del país.

Tal como sugieren estas limitaciones en cuanto a fabricación y suministros, los plazos requeridos para que las nuevas tecnologías tengan verdadero impacto en el ahorro de combustible son largos. Schafer et al. (2006) los han dividido en tres etapas, que se muestran en el cuadro 1.

En la primera etapa ha de desarrollarse una tecnología que sea competitiva dentro del mercado. Para ello debe estar accesible dentro de distintas categorías de vehículos y a un coste lo suficientemente bajo para que pueda implantarse a gran escala. El cuadro 1 calcula cuánto tiempo tardarían estos nuevos sistemas de propulsión de vehículos en convertirse en alternativas mayoritarias dentro del mercado. De éstas, sólo la gasolina con turbocargador, el híbrido de batería y gasolina y el diesel con catalizador están ya listas para empezar a producirse.

Aunque para los vehículos híbridos recargables no se ha anunciado todavía ningún plan relativo a su comercialización, varios fabricantes de automóviles han expresado interés en desarrollarlo dentro de la próxima década. La situación para vehículos de célula de combustible es más incierta. Un sondeo realizado entre los grandes fabricantes de automóviles sugiere que se trata de un proyecto que tardará al menos diez años en desarrollarse.

En la segunda etapa de implementación de la tecnología que muestra el cuadro, la de penetración en la producción actual de vehículos, se calcula el plazo que tardará esta tecnología en alcanzar una cuota de mercado que suponga un tercio de las cifras de venta de vehículos globales. En términos generales, estos plazos reflejan expectativas acerca de la viabilidad a gran escala de estos sistemas de propulsión basadas en las limitaciones existentes de ingeniería y de costes.

La tercera etapa de la implementación es el aumento significativo del uso real de dichos vehículos. Una reducción significativa de consumo de combustible sólo se produce cuando un gran número de vehículos de consumo eficiente se encuentran en circulación. Ello sucederá en un plazo de tiempo comparable a la vida media de un vehículo, que gira en torno a los 15 años.

En conjunto vemos que los plazos de tiempo globales antes de que el impacto de las nuevas tecnologías sea significativo son largos.

Impactos en el mundo real

La figura 7 muestra un ejemplo del hipotético impacto de una presencia creciente de vehículos eficientes en Estados Unidos. Este panorama ilustrativo asume que los volúmenes de producción de vehículos con motor de gasolina con turbocargador, de diesel con catalizador, de híbridos y de híbridos recargables aumentarán de forma continuada hasta alcanzar las cuotas de mercado reflejadas en la figura bajo el epígrafe de 2035. También se dan por hecho muchas otras cosas, claro está (Bandivadekar et al. 2008). La línea de «Ausencia de cambios tecnológicos» muestra que el consumo de gasolina crece de forma continuada debido al aumento de tamaño y kilometraje de los vehículos. Con el «énfasis en la reducción del consumo de combustible» (reflejado como ERFC en la figura) en un 50%, la mitad de las mejoras en cuanto a eficiencia son en términos de reducción de consumo real de combustible. La pequeña pero creciente porción dedicada a cada una de las nuevas y más eficientes tecnologías de sistemas de motor y propulsión está clara. En líneas generales este panorama reduce el consumo de combustible de 765.000 millones a 594.000 millones de litros por año, lo que supone un 22%. Las dos aportaciones más importantes que harían posible este panorama son el énfasis en la reducción del consumo de combustible y el porcentaje que ocuparía la tecnología más eficiente de todas — los coches híbridos— en el estimado de ventas para 2035.

Con un énfasis del 100% en la reducción del consumo de combustible en lugar del 50%, el consumo se reduce en un 15% adicional, a 505.000 millones de litros por año. Si el volumen de ventas de vehículos híbridos se duplica —es decir, si un 50% de los nuevos vehículos de 2035 son híbridos—, sería posible una reducción adicional en el consumo de combustible del 10%, es decir, estaríamos en 543.000 millones de litros al año. Combinadas, estas dos cifras darían una reducción adicional del 30%, como resultado de su implementación en el mercado estimada para 2035 en la figura 7. Nótese que el efecto de estos vehículos de tecnología eficiente crece lentamente al principio, pero a partir de 2030 desempeña un papel cada vez más importante a la hora de reducir el consumo de combustible y las emisiones de GEI, conforme las tecnologías avanzan y su penetración en el mercado crece.

Este panorama pone de manifiesto que los efectos inexorables del crecimiento del tráfico pueden ser contrarrestados, y que el consumo de combustible de emisiones de GEI puede congelarse y después reducirse. Pero lograr esto llevará aún un par de décadas. Acciones que afecten directamente a la flota de vehículos actualmente en uso, como reducir la circulación subiendo los precios del combustible, centrarse en la reducción de *consumo de combustible en la carretera* y controlar el aumento en el rendimiento y el tamaño de los vehículos son de suma importancia.

A modo de ilustración de los pasos necesarios para emprender esta senda, hemos analizado lo que costaría reducir a la mitad el consumo actual de combustible o, lo que es lo mismo, duplicar el ahorro de combustible de la nueva flota de vehículos que estará en el mercado en 2035 (Cheah et al. 2008). Sería necesario que dos terceras partes de las vehículos fabricados fueran híbridos, que el 75% de las mejoras en energía eficiente se centraran en la reducción real de consumo de combustible y no en aumentar el rendimiento y el tamaño de los vehículos (en Estados Unidos este tanto por ciento ha sido 0; en Europa ha sido de alrededor del 50%) y también una reducción de 20% en el peso de los vehículos. Aunque plausible, se trata de una tarea compleja, laboriosa y exigente.

Podríamos pensar que los objetivos fijados para 2020 referentes a una reducción de un tercio (por ejemplo, las regulaciones relativas a ahorro de combustible establecidas por el US CAFE, Corporate Average Fuel Economy, Promedio empresarial de ahorro de combustible) son más factibles, pero la realidad es bien distinta. A sólo 10 años del objetivo, las mejoras en consumo de combustible relacionadas con las nuevas tecnologías disponibles son menores, como menor es también el tiempo de que disponemos para aumentar su producción de manera significativa. Por tanto, los objetivos fijados para 2020 resultan ser tan poco factibles como los fijados para 2035.

Si miramos más hacia el futuro todavía, a 2050, vemos que aún nos encontramos lejos de lograr nuestro propósito: las emisiones de GEI serán únicamente entre

un 20-30% menores que hoy. Es aquí donde las tecnologías híbridas o de célula de combustible tendrán que hacer su entrada en el mercado a gran escala con sus distintas fuentes de energía, electricidad e hidrógeno. Si estos «portadores de energía» logran producirse con niveles realmente bajos de emisión de gases de efecto invernadero, entonces serán efectivos. Pero los plazos de tiempo para que cambios tecnológicos y de portadores de energía tan radicales tengan un impacto significativo son muy largos, de décadas, tal como sugiere el cuadro 1. Y el éxito del desarrollo de tecnología de batería y células de combustible, y de infraestructuras de baja producción y emisión de GEI es todavía incierto. En la actualidad se llevan a cabo intensas investigaciones centradas en estas opciones, y así debe ser.

Otras opciones

Hay otras opciones. Durante décadas se han estudiado mejoras en los sistemas de transporte y algunas de ellas han terminado por implementarse. Otras muchas no se han hecho por la dificultad que entraña coordinar empresas, gobiernos locales, regionales y nacionales, así como adaptar las nuevas tecnologías a las infraestructuras existentes.

En el transporte de pasajeros las oportunidades podrían ser significativas, porque el patrón actual de un pasajero por vehículo es ineficaz en términos de ahorro energético y económico. Muchos estudios demuestran el potencial para reducir en consumo de energía y las emisiones que tendría el llamado Sistema de Transporte Inteligente (o ITS por sus siglas en inglés), que pone en contacto electrónico a los usuarios de vehículos y reduce la circulación reorganizando patrones de desplazamiento, reforzando el transporte colectivo, sustituyendo los desplazamientos por tecnologías de información y comunicación y fomentando el uso de medios de transporte sin motor. Las opciones son muchas y atractivas, y ofrecen beneficios adicionales, tales como menor retraso ocasionado por los desplazamientos, menor inversión en infraestructura viaria y menor contaminación. Los beneficios energéticos y climáticos resultarían aún más evidentes a largo plazo.

El Sistema de Transporte Inteligente se basa en tecnologías que detectan los vehículos individuales y sus características, tales como velocidad o ubicación dentro de la red de transportes. Las tecnologías disponibles son varias: aparatos que detectan vehículos empleando una infraestructura especializada en carretera, el Global Positioning System o GPS o la red de teléfonos móviles. Para que esta información resulte de valor a un número mayor de usuarios, debe ser comunicada desde el vehículo a la infraestructura o a otros vehículos, para permitir así recabar datos sobre el estado general de la red. Recopilar esta cantidad ingente de datos y reducirlos a un formato que sirva para dar información a conductores individuales o para gestionar la red de transportes requiere detección, comunicación, computación, análisis y retroalimentación.

El Sistema de Transporte Inteligente trata de buscar soluciones a los problemas que en el futuro afectarán al rendimiento de nuestro actual sistema de transportes de superficie. La congestión, que es reflejo directo de la capacidad insuficiente de nuestras autopistas, constituye un problema clave que afecta a la circulación de personas y mercancías. Los costes —tantos financieros como ambientales— de solucionar estos problemas de capacidad construyendo infraestructuras nuevas o expandiendo las existentes pueden ser elevadísimos, en especial en áreas urbanas donde están vigentes las leyes que limitan el suelo urbanizable. Conceptos inspirados en el ITS pueden ayudar a combatir la congestión circulatoria de dos maneras: a través del uso de información sobre el tráfico, y a través de la aplicación de impuestos viarios. La seguridad en las autopistas es un motivo de preocupación añadido.

Este estudio se centra principalmente en el impacto ambiental y energético de los transportes. Reduciendo el flujo de la circulación, se reducen también el consumo de combustible y la tasa de emisión de GEI por vehículo. Sin embargo, si aumenta el uso de las infraestructuras debido a una mayor capacidad de éstas, el consumo de combustible y las emisiones también crecerán. Es por tanto importante explorar distintas aplicaciones del ITS para mejorar el sistema de transportes de modo que actúe en sinergia con el resto de las cuestiones enumeradas (Sussman 2005).

Lograr una reducción significativa en el uso de vehículos y en la emisión de GEI requiere una combinación de estrategias. Los cambios antes explicados referentes a la eficiencia del vehículo, y los aplicables al sistema de transportes para evitar la congestión circulatoria a menudo dan por sentado que los patrones residenciales y de desplaza-

miento continuarán como en la actualidad. Si los patrones de urbanización cambian y conducen a una mayor densidad demográfica en áreas metropolitanas en detrimento de la población suburbana y rural, entonces se reduciría la circulación por carretera. Si las personas están dispuestas a vivir en corredores urbanos más poblados, podrían beneficiarse más del transporte público. Y entonces los «minicoches» urbanos cobrarían protagonismo. Sin embargo, los patrones de urbanización se han desplazado durante más de seis décadas precisamente en dirección contraria: hacia las áreas suburbanas.

Para que puedan desarrollarse sistemas de transporte más diversificados sería necesario implantar dos tipos de políticas: controlar el negocio urbanístico y subir los impuestos por el uso de carreteras. Son opciones que deberían estudiarse seriamente. Las ventajas de un esfuerzo concentrado por mitigar la congestión circulatoria y su impacto ambiental, por limitar la circulación de vehículos con un solo pasajero y fomentar el uso de automóviles ecológicos y de pequeño tamaño dentro de la ciudad serían muchas y significativas.

¿Qué cabe esperar del futuro?

La figura 8 muestra los objetivos en cuanto a tasa de emisión de gases de efecto invernadero de vehículos ligeros en Europa y Estados Unidos. Se trata de objetivos muy agresivos considerando lo breve de los plazos establecidos, y por tanto requieren un ritmo de progreso más rápido que el factor de reducción 2 fijado para dentro de 25 años. En Europa conseguir estos objetivos a corto plazo será especialmente difícil porque el margen de actuación es menor con la flota de vehículos en uso: la mitad de la misma ya es diesel, y el tamaño y el peso de los vehículos es aproximadamente un tercio del de los estadounidenses. Asimismo, el aumento del rendimiento de los coches en Europa y Japón, que ya se ha producido, ha sido significativamente más bajo que en Estados Unidos. Las oportunidades para futuras mejoras son, por tanto, menores.

La figura 9 ilustra nuestro desafío. Muestra las proyecciones globales y de Estados Unidos para el año 2050 en cuanto a emisión de GEI en vehículos ligeros. Hoy la flota estadounidense es responsable de cerca de la mitad de las emisiones globales de estos gases dentro del sector de movilidad. Europa y Japón aportan una fracción significativa de la mitad restante. Incluso con una implementación agresiva de tecnología más eficiente, la reducción de la tasa de emisiones es muy modesta.

En resumen vemos que el potencial de una reducción en el consumo de combustible de vehículos ligeros —coches, todoterrenos y camiones ligeros— durante los próximos 10 a 25 años está entre en 30 y el 50%. Vendrá de las mejoras en los motores, del desarrollo de trenes de transmisión más eficientes y de la reducción del peso y el tamaño de los vehículos. Que alcancemos o no este objetivo estará en función de cómo controlemos nuestras expectativas res-

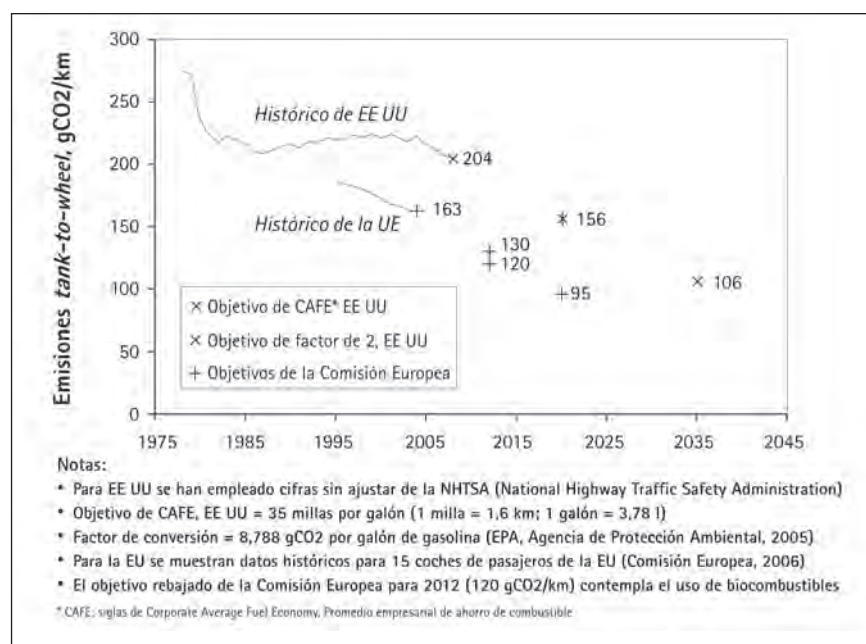


Figura 8. Objetivos y regulaciones referentes a tasa de emisión de gases de efecto invernadero en Estados Unidos y Europa.

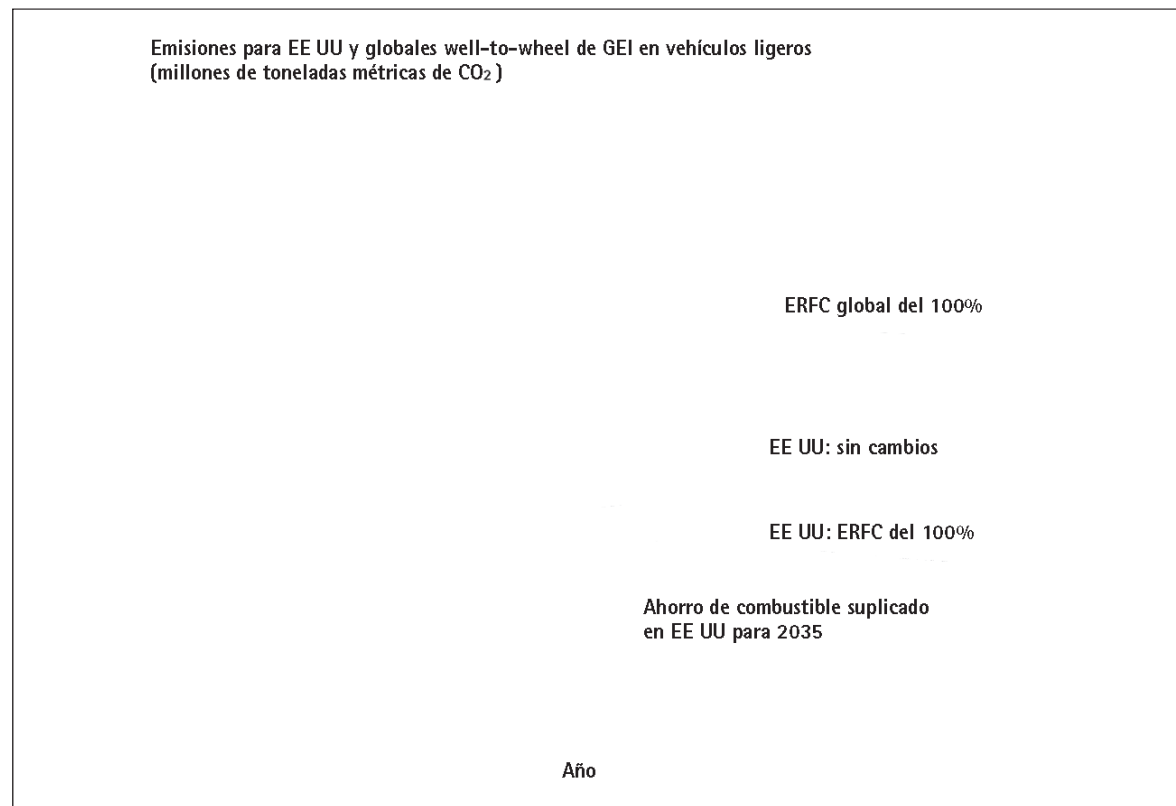


Figura 9. Panorama global y de Estados Unidos para 2050 en el que se muestran los efectos de reducir a la mitad el consumo de combustible en vehículos nuevos para 2035 y las emisiones de GEI *well-to-wheel* en vehículos en uso. El rendimiento y el tamaño medio de los vehículos son constantes (énfasis del 100% en reducir el consumo de combustible) (Bandivadekar et al. 2008).

pecto al rendimiento de nuestros vehículos y sobre todo de cuán urgentemente acometamos las reformas tecnológicas necesarias. Esto último dependerá de que el contexto sea propicio: el precio del petróleo, nuestra percepción acerca de su disponibilidad, el grado de concienciación respecto a las emisiones de gases de efecto invernadero, la exhaustividad y eficiencia de las medidas y regulaciones impuestas y nuestra disposición a moderar nuestro grado de movilidad. A corto plazo, en los 10 o 20 años siguientes, la prioridad deberá ser reducir el consumo de petróleo destinado al transporte. A medio plazo, puesto que deberíamos concentrar cada vez más esfuerzos en reducir la emisión de GEI, el camino resulta menos claro en tanto que la necesidad de seguir reduciendo la producción de petróleo y las emisiones se vuelve más perentoria. Entonces habremos de resolver las cuestiones de cuánta contribución podremos obtener de los biocombustibles, el grado hasta el cual podremos confiar en la electricidad como principal fuente de energía para nuestros vehículos y si el hidrógeno y la célula de combustible deberían ser nuestro objetivo último. A día de hoy no tenemos capacidad de resolver dichas cuestiones, pero necesitamos desarrollar los conocimientos básicos de modo que, con el tiempo, podamos dar respuestas mejores y más precisas.

La demanda mundial de transporte crece de forma inexorable y no existe un avance que por sí solo pueda resolver los crecientes problemas de consumo de com-

bustible y emisión de GEI. Este ensayo ha explicado que el progreso debe provenir de un esfuerzo exhaustivo por desarrollar y comercializar vehículos más eficientes y combustibles más benignos con el medioambiente, y a animarnos a todos a usar nuestros vehículos y otros medios de transporte de modo que ahorremos energía. Todos estos cambios habrán de ser implementados a gran escala para que sus efectos en cuanto a consumo de petróleo destinado al transporte, uso energético y emisiones de GEI sean significativos. Es probable que dicha implementación aumente el precio que los usuarios han de pagar por los transportes, y que requiera medidas gubernamentales que fomenten e incluso obliguen a encaminarse hacia estos objetivos, compartiendo el peso que ello supone de forma equitativa y minimizando los costes sociales.

La transición de nuestra situación actual a una en la que el consumo de petróleo y la emisión de gases sean menores, incluso en el mundo desarrollado, llevará varias décadas, más de las que imaginamos. No debemos olvidar que lo fundamental es llevar a cabo cambios que contribuyan de manera significativa a conseguir los objetivos. Necesitaremos mejores tecnologías, vehículos más apropiados, fuentes de combustible más «verdes» y, también, tomar conciencia de la necesidad de «conservar», de emprender la senda de la sostenibilidad. Y todo ello deberemos hacerlo con urgencia y determinación.

Bibliografía

- Bandivadekar, A., K. Bodek, L. Cheah, C. Evans, T. Groode, J. Heywood, E. Kasseris, M. Kromer y M. Weiss. «On the Road in 2035: Reducing Transportation's Petroleum Consumption and GHG Emissions». MIT Laboratory for Energy and the Environment, Massachusetts Institute of Technology, julio de 2008. <http://web.mit.edu/sloan-auto-lab/research/beforeh2/otr2035/>.
- Cheah, L., A. Bandivadekar, K. Bodek, E. Kasseris y J. Heywood. «The Trade-off between Automobile Acceleration Performance, Weight, and Fuel Consumption». *Society of Automotive Engineers*, artículo 1 (2008): 1.524, SAE International Powertrains, Fuels and Lubricants Congress, Shangai, China, 23-25 de junio de 2008.
- Intergovernmental Panel on Climate Change. *Climate Change 2007: Synthesis Report. Summary for Policy Makers*, Reino Unido: Cambridge University Press, 2007.
- Kasseris, E. y J. B. Heywood. «Comparative Analysis of Automotive Powertrain Choices for the Next 25 Years». *Society of Automotive Engineers*, SAE paper 1 (2007): 1.607.
- National Research Council. «Toward a Sustainable Future: Addressing the Long-Term Effects of Motor Vehicle Transportation on Climate and Ecology». Transportation Research Board, Washington D.C.: National Academy Press, 1997.
- Schafer, A., J. Heywood y M. Weiss. «Future Fuel Cell and Internal Combustion Engine automobile Technologies: A 25-Year Life Cycle and Fleet Impact Assessment». *Energy*, vol. 31, Nº. 12 (2006): 2.064-2.087.
- Sussman, J. *Perspectives on Intelligent Transportation Systems (ITS)*. Nueva York: Springer, 2005.
- World Business Council for Sustainable Development. «Mobility 2001: World Mobility at the End of the Twentieth Century and its Sustainability». Informe preparado para el Sustainable Mobility Working Group (Grupo de trabajo sobre movilidad sostenible, WBCSD), elaborado por Massachusetts Institute of Technology y Charles River Associates. Disponible en Internet en: <http://www.wbcdsmobility.org>, agosto de 2001.
- , «Mobility 2030: Meeting the Challenges to Sustainability». World Business Council on Sustainable Development. Informe sobre el proyecto de movilidad sostenible, 180 páginas. Disponible en Internet en: <http://www.wbcd.org/DocRoot/fl311MAvneJpUcnLgSeN/mobility-full/pdf>, 2004.

retos actuales de la energía

CAYETANO LÓPEZ

Introducción

La creciente demanda mundial de energía y los problemas, tanto de escasez como de impacto ambiental, asociados a las fuentes convencionales están en la base de una muy probable crisis energética en las próximas dos o tres décadas. En efecto, el petróleo será cada vez más caro y escaso, mientras que los efectos climáticos de la utilización masiva del conjunto de las fuentes fósiles serán ya evidentes. Al tiempo, la vida útil del parque nuclear actual habrá llegado a su fin y no está claro, en particular en Europa, si la potencia instalada que dejará entonces de estar operativa se reemplazará con nuevos dispositivos nucleares.

En este horizonte temporal no se puede prescindir de ninguna de las fuentes de energía existentes, con las modificaciones que sean imprescindibles para eliminar o amortiguar sus efectos ambientales, y hay que incorporar nuevas fuentes, especialmente las renovables. En lo que sigue, haré una descripción del estado de las tecnologías a nuestra disposición y los desarrollos más prometedores en cada una de ellas, siempre en la perspectiva temporal de unas pocas décadas.

A más largo plazo, la fusión nuclear será parte de un catálogo más sostenible de fuentes energéticas, pero no estará lista en el periodo de tiempo considerado y, por tanto, no podrá contribuir a solventar la crisis. Ésa es la razón de que no dedique un apartado a la fusión nuclear,

que es un potente e interesante programa que se desarrolla a escala mundial y cuyo objetivo es el dominio de las reacciones de fusión nuclear como fuente de energía, pero cuyos ritmos previsibles la sitúan fuera del espacio temporal elegido para analizar los problemas energéticos.

Energía y civilización

La energía es un ingrediente fundamental en la vida de la humanidad. No hay proceso industrial, agrícola, sanitario, doméstico o de cualquier otro tipo que no requiera un aporte de energía exosomática. Los seres humanos consumimos, a través de los alimentos, unas 2.500 kilocalorías por día, mientras que, en los países industrializados, la cantidad de energía suplementaria (exosomática) que se consume, por término medio, en el conjunto de las actividades humanas –industriales, domésticas, transporte u otras– es el equivalente de 125.000 kilocalorías por persona y día, es decir cincuenta veces más, llegando en el caso de Estados Unidos al doble de esta cifra (véase, por ejemplo, British Petroleum 2008). De hecho hay una fuerte correlación entre el consumo de energía unitario y la prosperidad de las distintas sociedades.

Si se representa, como en la figura 1, cada país en un diagrama en el que el eje Y especifica el Índice de Desarrollo Humano (IDH) para ese país, tal como lo determina la ONU a partir de los datos básicos de bienestar de sus habi-

tantes, y el eje X el consumo anual de energía per capita (en este caso en forma de electricidad), se observan dos fenómenos interesantes. En los países más pobres, la correlación es muy fuerte, suponiendo el aumento del consumo energético ganancias evidentes en su IDH, mientras que para los países más desarrollados las diferencias en el consumo energético, que son muy grandes, no repercuten de forma significativa en su nivel de bienestar. Lo que implica que para estos últimos el ahorro energético es una política posible y deseable; en realidad el ahorro es la fuente de energía más limpia y abundante en el grupo de los países más prósperos. Por el contrario, el necesario desarrollo económico y social de los países comparativamente pobres, que suponen la parte mayoritaria de la población mundial, requerirá inevitablemente de un mayor consumo de energía, con lo que es poco realista pensar que el consumo global de energía va a disminuir en el futuro. Si lo hiciera, supondría una verdadera catástrofe para los países menos desarrollados, que tienen escasez de todo, también de energía. Así, aunque las políticas activas deban ir encaminadas en los países del primer mundo de forma preferente al ahorro energético, desde el punto de vista global hay que enfrentar el problema de una demanda energética creciente.

Las fuentes de energía actuales

Las fuentes de energía primaria están identificadas y no parece que su catálogo pueda ampliarse en un futuro previsible. Desde los albores de la humanidad hasta el comienzo de la era industrial, a principios del siglo XIX, las únicas fuentes de energía primaria disponibles eran la leña

y otras formas de biomasa natural, los animales de carga y el viento para el transporte marítimo o fluvial. A partir de la aparición de las primeras máquinas de vapor empieza a utilizarse como fuente de energía el carbón, que continúa hoy suponiendo una parte importante de la energía primaria consumida, y más tarde, cuando se extiende el uso del automóvil dotado de un motor de combustión interna que debe ser alimentado con un fluido combustible, el petróleo y sus derivados pasan a ocupar un lugar preeminente como fuente de energía. Finalmente, el gas natural ha ido configurándose a lo largo del último medio siglo como un componente importante en la generación de electricidad y en la producción de calor para usos industriales y domésticos.

Estos combustibles, carbón, petróleo y gas natural, se encuentran a diferentes profundidades en la corteza terrestre y se han formado en procesos a alta presión y temperatura a partir de materia orgánica, principalmente plantas y organismos marinos, de épocas geológicas pasadas. De ahí que se les conozca como combustibles fósiles. Su contribución al conjunto de la energía primaria consumida en el mundo a finales de 2007 (British Petroleum 2008) es del 35,6% el petróleo, 28,6% el carbón y 23,8% el gas natural; en conjunto representan, por tanto, el 88% del total, lo que por muchas razones, como veremos en lo que sigue, configura un escenario insostenible incluso a corto plazo. El resto proviene de la energía nuclear, que supone un 5,6% del total, y de las energías renovables, esencialmente la hidroeléctrica. Las energías procedentes del viento y del sol en sus diversas modalidades son hoy un factor marginal desde el punto de vista global, aunque



Figura 1. El Índice de Desarrollo Humano (IDH) como función de la energía eléctrica consumida por persona y año. Elaboración propia a partir de los datos del IDH elaborados por la ONU (UN 2006) y de los datos sobre consumo eléctrico de la Agencia Internacional de la Energía (IEA 2008).

empiezan a tener una mayor presencia en algunos países, singularmente en España. Y éste es el panorama global; no hay más fuentes disponibles de energía primaria.

De esta energía primaria total, una parte importante se transforma en electricidad –en un país como España aproximadamente el 40%– mientras que el resto se dedica al sector del transporte y otros usos industriales o domésticos.

Los combustibles fósiles

La enorme predominancia de los combustibles fósiles como fuente de energía primaria tiene algunas consecuencias importantes:

En primer lugar, están repartidos de forma desigual. Las dos terceras partes de todas las reservas conocidas de petróleo, que es probablemente el combustible más difícil de sustituir, se encuentran en el subsuelo de cinco o seis países del Medio Oriente, lo que implica un grado de dependencia poco compatible con la estabilidad del suministro. El gas natural también está muy concentrado en esa zona y en los países de la antigua URSS, mientras que el carbón está más repartido por todas las regiones del planeta.

En segundo lugar, son materias primas no renovables, debido a que se han formado a lo largo de decenas o cientos de millones de años y, por lo tanto, su reposición es imposible. Por otra parte son recursos limitados. En particular, la utilización del petróleo como fuente de energía sobre la que se basa el modo de vida de los países industrializados, podría ser una breve fluctuación en la historia de la humanidad, restringida a un periodo del orden de un par de siglos.

En tercer lugar, se trata de materias primas escasas. Sobre la cantidad disponible de petróleo hay una cierta discusión, pero la mayoría de los geólogos y expertos petroleros coinciden en que, al ritmo de consumo actual, que es nada menos que unos 85 millones de barriles de petróleo diarios –lo que equivale a quemar mil barriles de petróleo por segundo–, tenemos para unos pocos decenios. Se puede argüir que la cantidad de petróleo extraído dependerá del precio y que, si éste sube, no habrá límite práctico a su producción. Pero ese argumento obvia el hecho de que cuesta cada vez más energía –en prospecciones, bombeos, tratamiento y logística– extraer petróleo de yacimientos cada vez más profundos o agotados. A mediados del siglo xx, para extraer un barril de petróleo, bastaba una fracción del orden del 1% de la energía contenida en ese mismo barril; hoy esa fracción está entre el 10 y el 15%. Cuando el gasto en energía necesario para extraer un barril de crudo se acerque al generado por la combustión de su contenido, no importará su precio: habrá desaparecido como fuente de energía primaria aunque sea útil a otros fines, especialmente en la industria petroquímica para la síntesis de multitud de compuestos básicos en casi todas las ramas de la industria y la agricultura. Las reservas probadas de petróleo, consumidas al ritmo actual dan para unos cuarenta años, mientras que

las de gas llegan a los sesenta años y las de carbón hasta un siglo y medio, aproximadamente (British Petroleum 2008). Por supuesto habrá nuevos descubrimientos y los llamados petróleos no convencionales a partir de hidrocarburos dispersos en arenas o esquistos bituminosos o en alquitranes pesados, pero hay que tener siempre en cuenta su creciente coste energético y, por lo tanto, su decreciente rendimiento neto y mayor precio. En todo caso, no se producirá una brusca discontinuidad en el suministro pasando del consumo actual a nada. Lo probable es que se produzca un aumento progresivo de precio y, a partir de cierto momento, una disminución también progresiva de consumo y producción.

Por último, sabemos que quemar combustibles fósiles supone la emisión de enormes cantidades de dióxido de carbono (CO₂) a la atmósfera. Este gas es uno de los que producen el efecto invernadero y, por lo tanto, contribuye al calentamiento global del planeta que, dada la rapidez del fenómeno en términos geológicos, podría desencadenar perturbaciones climáticas graves potencialmente dañinas para nuestra civilización –no para la vida, como con frecuencia se dice, ni para la vida humana, pero sí para nuestra compleja y exigente organización social.

En definitiva, nuestra actividad social se basa en el consumo de combustibles fósiles que, por motivos de escasez y medioambientales, debe limitarse en el futuro. A pesar de lo cual, el carbón seguirá siendo durante decenios una fuente de energía masiva, pero su uso sólo será tolerable si se palia la contaminación que produce.

En consecuencia, el segundo reto energético –el primero es reducir el consumo en los países desarrollados– es disminuir la contribución de los combustibles fósiles en nuestro menú.

Preparar reemplazos al petróleo

El sector del transporte depende casi en su totalidad de los combustibles líquidos derivados del petróleo. El carbón y el gas natural son hoy importantes en la producción de electricidad pero es concebible su sustitución por energías renovables o energía nuclear a largo plazo, mientras que no es fácil imaginar alternativas al uso de los derivados del petróleo en el transporte. Todas ellas pasan por cambios de enorme envergadura.

La primera alternativa posible es el uso de biocarburantes, bioetanol y biodiesel, que sustituyan al menos una parte de los carburantes convencionales. Pero muy recientemente hemos tenido oportunidad de observar los problemas colaterales que pueden surgir, especialmente en el sector de la alimentación, aun cuando la producción de biocarburantes está todavía en un estadio muy incipiente. Desde luego la influencia de la producción de bioetanol, por poner el caso más controvertido, en el aumento de los precios de los alimentos es limitada y concurre con otras causas de mayor calado, ya sean coyunturales o estructurales. El único cereal que se emplea de forma significa-

tiva en la fabricación de etanol es el maíz, mientras que el trigo y la cebada se emplea en cantidades marginales respecto de la producción total y el arroz no se emplea en absoluto. Sin embargo, los precios han aumentado en todos los casos, y de forma prominente en el caso del arroz. Por otra parte, del orden de la mitad del bioetanol producido actualmente proviene de la caña de azúcar brasileña, y no ha habido aumentos en el precio del azúcar.

En todo caso, el etanol fabricado a partir de cereales es la peor solución posible, no sólo por razones de impacto sobre la alimentación, sino especialmente por su escaso rendimiento energético. En efecto, la cantidad de energía contenida en un litro de etanol es apenas superior a la que es necesaria gastar para obtenerlo a partir de cereales, en fertilizantes, semillas, cosechas, transporte y tratamiento (véase, por ejemplo, Worldwatch 2007 o CIEMAT 2005). Por lo tanto, desde el punto de vista energético no es razonable utilizar este tipo de materia prima. Hay además razones de tipo medioambiental relacionadas con el uso del agua y de tierra cultivable que inciden en esta misma dirección (Zah 2007). Por el contrario, el rendimiento energético de la caña de azúcar es muy superior, y todavía lo es más el etanol a partir de lo que se llama biomasa lignocelulósica, presente en plantas leñosas o herbáceas y en residuos orgánicos. Es lo que se llama etanol de segunda generación. Todas estas conclusiones se encuentran recogidas en el interesante gráfico de la figura 2, reproducida de Zah 2007, en la que se recogen todos los datos sobre consumo de energía fósil en las labores de cultivo, cose-

cha, pretratamiento y demás procesos necesarios para obtener biocarburantes de diferentes materias vegetales, así como el conjunto de los impactos ambientales, en relación con la utilización directa de los derivados del petróleo.

El tercer reto, por tanto, es perfeccionar la tecnología ya existente en la producción de biocarburantes de segunda generación para llevarla a un nivel de explotación industrial. No se está muy lejos y ya hay algunas plantas piloto en las que se están experimentando varios procesos de fabricación de etanol a partir de este tipo de biomasa que no tiene efectos sobre la alimentación, incurre en menos costes energéticos y tiene menos contraindicaciones ambientales (véase, por ejemplo, Ballesteros 2007 o Signes 2008). Así, el etanol de caña y los biocombustibles de segunda generación podrían disminuir la dependencia del petróleo en el sector del transporte, aunque no podrían suprimirla debido a las limitaciones en tierra cultivable y biomasa disponible en comparación con los gigantescos consumos del sector.

Es más fácil, al menos en principio, sustituir los combustibles fósiles empleados en la generación de electricidad por otras fuentes, renovables o nuclear, que sustituir de forma completa los derivados del petróleo, por lo que, a más largo plazo pienso que se irá hacia vehículos eléctricos, en una primera etapa en la modalidad de híbridos y más adelante sólo eléctricos. El problema aquí es el almacenamiento de electricidad, poco eficiente y muy contaminante, que se realiza en las baterías utilizadas actualmente. Pero hay un intenso trabajo de investigación en nuevos dispositivos de almacenamiento de electricidad

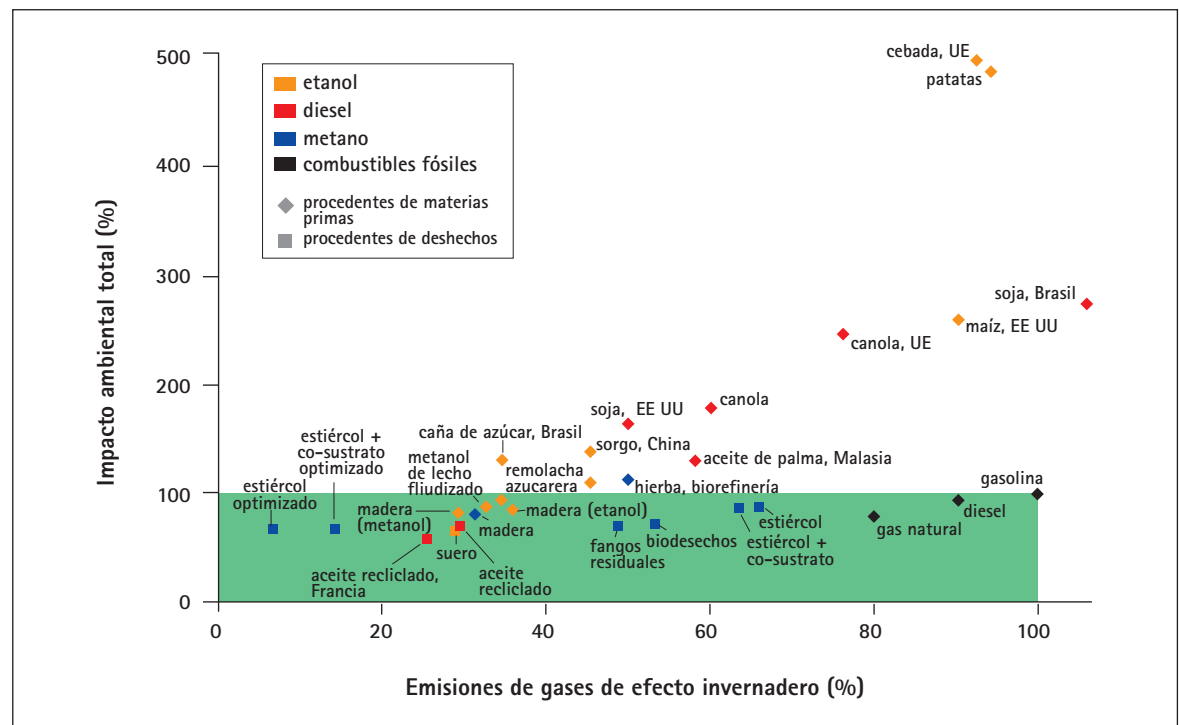


Figura 2. Impacto ambiental y consumo de energía fósil en la producción de biocarburantes relativos a la utilización directa de carburantes derivados del petróleo (Zah 2007).

	Carbón	Gas Natural	Nuclear	Hidráulica	Otras Renovables	Otros
Promedio mundial (06)	40%	20%	16%	16%	2%	6%
EEUU (06)	49%	20%	19%	7%	2%	3%
Francia (06)	4%	4%	78%	11%	1%	2%
China (04)	83%	0%	2%	15%	0%	0%
España (07)	24%	34%	18%	10%	11%	3%

Tabla 1. Generación de electricidad por fuentes de energía primaria, en porcentaje del total.

que permitirán la construcción de automóviles eléctricos con las prestaciones adecuadas.

En general, debe decirse que el problema del almacenamiento de energía, sea en forma de electricidad, calor, hidrógeno o cualquier otra, está ocupando una posición central en la investigación energética, tanto por su importancia para el futuro del sector del transporte como para solventar los problemas derivados de la intermitencia de las fuentes renovables, como veremos más adelante. Es decir, si se consigue avanzar en la tecnología de almacenamiento de electricidad (véase, por ejemplo, el trabajo de José Luis Mata Vigil-Escalera en Club de la Energía 2008a), un reto formidable si se quieren reproducir las prestaciones del coche de gasolina, el futuro previsible de una parte importante de los vehículos será eléctrico. Me concentraré, por tanto, en lo que sigue en la producción de electricidad, que se configura como la forma de energía más flexible y adaptable incluso para el futuro del sector del transporte.

¿Carbón limpio?

El esquema de producción de electricidad varía mucho de unos países a otros. En la tabla 1 se dan algunos datos sobre la composición relativa de las fuentes de energía empleadas en la generación eléctrica en España, en algunos otros países y el promedio mundial (IEA Statistics y Club Español de la Energía 2008).

Puede verse que, con la excepción de Francia, que se basa de forma muy destacada en su parque nuclear, y parcialmente de España, con una apreciable contribución de las renovables, la fuente de energía básica sigue siendo el carbón. Y seguirá siéndolo durante mucho tiempo debido a su abundancia y a la distribución en prácticamente todos los continentes. Nótese, en particular, el caso de China, que, según la Agencia Internacional de la Energía, lleva algunos años poniendo en funcionamiento, en promedio, una nueva gran planta eléctrica de carbón por semana. Pero el carbón es, con diferencia, el combustible fósil más contaminante, ya que emite cerca del doble de dióxido de carbono a la atmósfera por unidad de energía producida que el gas natural y aproximadamente un 40% más que la gasolina usada en los motores de explosión, además de compuestos de azufre, nitrógeno y metales pesados.

Así que, si queremos seguir utilizando el carbón como fuente de energía debemos desarrollar procedimientos que eliminen o limiten las emisiones asociadas de CO₂ a la

atmósfera –las otras emisiones ya se controlan en las plantas de generación actuales–. Es lo que se conoce como procesos de captura y secuestro de CO₂ (CCS, de sus siglas en inglés). Este tipo de procesos se encuentra todavía en una fase muy preliminar. En particular, la captura del CO₂ emitido en la combustión del carbón puede hacerse con técnicas de oxidación, modificando la composición del aire que entra en las calderas de forma que los gases emitidos sean CO₂ prácticamente en su totalidad y no haya necesidad de separarlo, o bien con técnicas de separación tras la combustión con aire. En ambos casos se generarán costes energéticos adicionales y requerirán de nuevos procesos físico-químicos probados ya en el laboratorio pero no a la escala industrial requerida. En cuanto al secuestro, es preciso encontrar depósitos subterráneos o submarinos en los que inyectar el CO₂ con un grado de estanqueidad suficiente como para que dicho gas esté confinado durante siglos.

En realidad, depósitos de este tipo existen en la naturaleza. Por ejemplo, los yacimientos de gas natural que lo han mantenido confinado durante periodos geológicos, una vez explotados pueden ser utilizados para almacenar dióxido de carbono, lo mismo que yacimientos agotados de petróleo, formaciones sedimentarias salinas, etc. De hecho, la mayoría de las experiencias de secuestro de CO₂ en el mundo están asociadas a campos petrolíferos en producción ya decreciente, en los que se inyecta dicho gas a presión con objeto de mejorar sus prestaciones y obtener crudo que no brotaría con las tecnologías convencionales de extracción.

Otra experiencia interesante es la de Sleipner, un campo de producción de gas en la costa noruega del Mar del Norte. En este campo, el metano, que es el principal ingrediente del gas natural, sale mezclado con importantes cantidades de CO₂ que, una vez separado en la planta de extracción, vuelve a inyectarse a unos mil metros de profundidad en un lecho de rocas porosas con agua y sales. El secuestro de este CO₂ viene haciéndose desde el año 1996 y los datos

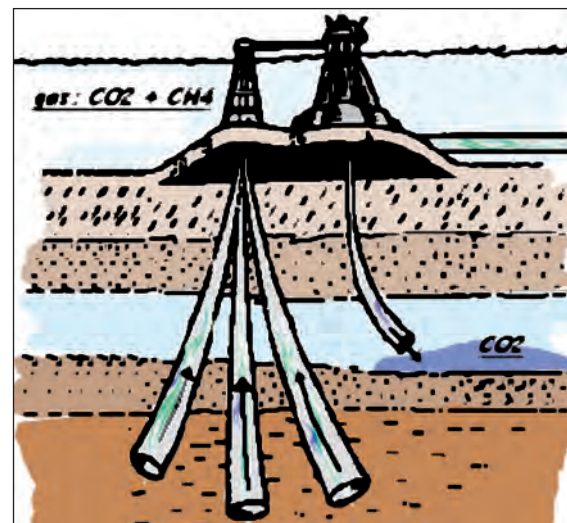


Figura 3. Campo de Sleipner, en la costa noruega del Mar del Norte.

sobre estanqueidad del depósito serán de gran valor a la hora de localizar y utilizar nuevos emplazamientos de forma masiva. En todo caso, hay que remarcar que los procesos de captura y secuestro de dióxido de carbono siempre supondrán un gasto adicional y, por tanto, encarecerán el precio de la energía obtenida a partir de uso limpio del carbón. Los expertos cifran este encarecimiento entre un 30 y un 100% del precio asociado al carbón sin CCS (Socolow 2005, Fundación para Estudios sobre la Energía 2008). De todas formas, este sobrecoste debe ser situado en un contexto de encarecimiento de las energías convencionales y renovables, de costes adicionales por emisiones de CO₂ y de ayudas, del tipo de las contempladas en el Régimen Especial en nuestro país, a las energías no contaminantes. La conclusión es que la humanidad no dejará de utilizar una fuente de energía tan abundante y repartida como el carbón, pero su uso tiene consecuencias ambientales graves por lo que es de la mayor importancia encontrar una forma de contrarrestarlas, por ejemplo con técnicas de CCS.

Electricidad renovable. El viento

Quizá el desafío más importante que tenemos para las próximas décadas es aumentar de forma significativa la contribución de las energías renovables desde el nivel

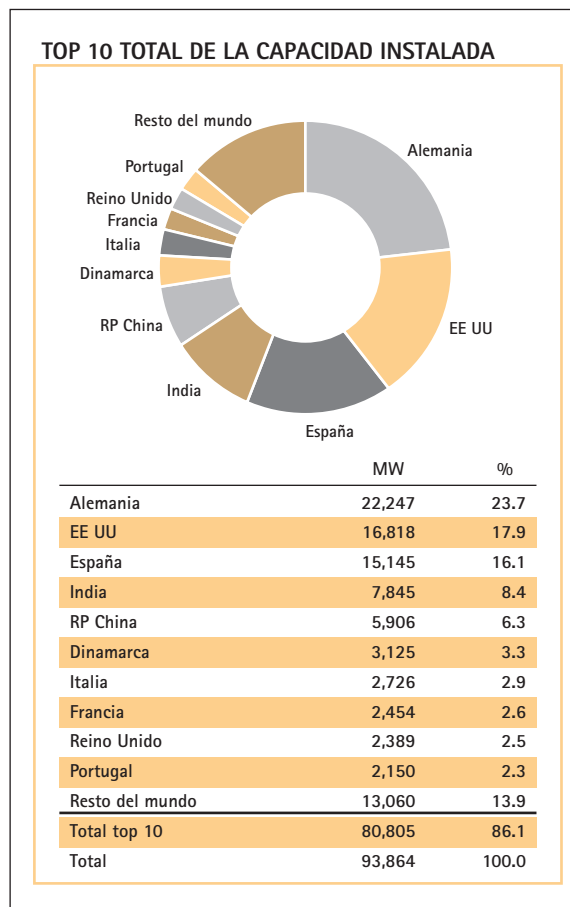


Figura 4. Potencia eólica instalada a 31 de diciembre de 2007.

actual, marginal en términos globales. Las energías renovables distintas de la hidráulica, que es la que tiene una mayor presencia y cuyos recursos han sido utilizados de forma más completa, es decir, esencialmente la eólica y la solar, tienen ventajas e inconvenientes. Las ventajas, que son el negativo de los inconvenientes que enumerábamos al hablar de los combustibles fósiles, derivan de su carácter sostenible, ilimitado, muy poco contaminante, incluso si tenemos en cuenta el ciclo completo de vida, y su distribución territorial. Los inconvenientes pueden agruparse en dos categorías: elevado coste e intermitencia.

Una de las razones del coste elevado de la electricidad renovable se debe a la dispersión de este tipo de energías, que es una característica intrínseca sólo compensada por su carácter ilimitado y sostenible, aunque es verosímil que los costes de las energías convencionales vayan aumentando y convergiendo con las renovables a medida que su disponibilidad disminuya y se imputen los costes de los efectos ambientales. Otra parte del coste elevado de las renovables se debe a que las tecnologías asociadas están todavía en un estadio poco avanzado de desarrollo. En este sentido, para disminuir los costes que derivan de la falta de desarrollo tecnológico es preciso construir un mercado de dimensiones globales. Sólo así se abaratará la producción de componentes, tanto por el aumento de las series producidas como por la aparición de más actores que acaben con el oligopolio existente en algunos de los campos clave, al tiempo que podrán implementarse mejoras en la operación y mantenimiento de las plantas renovables, tras un cierto periodo de experiencia operativa en condiciones de explotación industrial. Justamente, los diferentes sistemas puestos en funcionamiento para estimular el despliegue de las energías renovables, están dirigidos a ampliar ese mercado mediante subsidios u otro tipo de ayudas que compensen las dificultades iniciales.

Como es bien sabido, en España, como en algunos otros países avanzados en este campo, se ha optado por la conformación de un régimen especial de tarifas para las energías renovables (y la cogeneración), excluyendo la gran hidráulica. Se trata de una serie de incentivos, o subsidios, al kilovatio hora de origen renovable con objeto de compensar los mayores costes actuales y estimular así el crecimiento del sector. Las tarifas especiales son distintas para cada tecnología de generación, reflejando los diferentes costes actuales, y tienen vocación de ir disminuyendo en el tiempo según vayan afinándose los costes y convergiendo con los de las energías convencionales. Este o cualquier otro de los esquemas existentes se ha mostrado ya eficaz en la primera de las energías renovables que se puede considerar extendida en un mercado de dimensiones globales, la eólica. En efecto, a finales de 2007, ya existía una potencia eólica acumulada en el mundo de 93.900 MW (Global Wind Energy Council 2008), lo que ha permitido configurar un sector industrial dinámico y en crecimiento en todas las regiones del mundo.

Como se ve en la figura 4, los tres países que cuentan con una mayor potencia instalada son Alemania, Estados Unidos y España, aunque de los tres, y debido a su menor consumo total, es España quien obtiene una mayor fracción de su electricidad de esta fuente de energía, aproximadamente un 9%. De hecho, España ocupa el segundo lugar del mundo, tras Dinamarca, tanto en fracción de electricidad total de origen eólico como en potencia instalada por habitante (European Wind Energy Association 2008).

Los objetivos perseguidos con los estímulos públicos a la creación de un mercado eólico global están dando sus frutos, no sólo en la creación de nueva actividad industrial y empleo, sino también en el abaratamiento progresivo de la energía así producida. A finales de los setenta, con aerogeneradores del orden de 50 kW de potencia y un diámetro de rotor de unos 15 m, el precio unitario era del orden de 30 a 40 c€/kWh mientras que actualmente, con aerogeneradores de unos 2 MW de potencia, con un diámetro de rotor de cerca de 100 m, el coste de producción es sólo ligeramente más alto que el convencional, estando situada la tarifa de régimen especial para la electricidad eólica en un sobrepeso del orden de 2,9 c€/kWh (unos 2c\$/kWh en Estados Unidos).

Además del tamaño, se han producido muchas otras mejoras tecnológicas que afectan los elementos móviles, los materiales de que están contruidos, los sistemas de conversión, transformación y evacuación, y los procesos de fabricación y montaje. El reto en este campo consiste en proseguir la extensión del mercado y las mejoras tecnológicas que acabarán por hacer que el coste unitario de la electricidad producida iguale al de la convencional. Y conquistar el medio marino, con la llamada eólica *off-shore* debido a la mayor calidad del recurso eólico –vientos sostenidos y sin turbulencias–, aunque son considerables las dificultades para el anclaje y mantenimiento de los aero-

generadores cuando la lámina de agua no es muy delgada, así como de evacuación de la electricidad producida.

Tiene, por tanto, la energía eólica un largo recorrido en cuanto a su tecnología y a la extensión territorial a otros ámbitos, el marino desde luego, pero también el de la pequeña eólica, tanto en el entorno urbano como en el de las concentraciones humanas aisladas de la red o con una red débil. Como ocurre con todas las renovables, está por resolver el problema de la intermitencia. El viento es discontinuo, de forma que, para el caso de España, por ejemplo, los parques eólicos sólo generan energía durante unas dos mil horas al año en promedio, como puede verse en la figura 6, algo menos de la cuarta parte de las que componen un año.

Y no siempre se adecua el momento de la generación con el de los máximos de la demanda. No obstante, durante el mes de marzo de 2008, un mes ventoso, la energía eólica supuso nada menos que el 18,7% de la electricidad generada en España en ese mes y el día 22 hacia las 18h estuvieron activos 9.900 MW eólicos, lo que supuso un 41% de la electricidad global generada en ese momento, y durante todo el fin de semana del 21 al 23 de marzo, la electricidad de origen eólico representó un 28% de la producción total.

Para resolver el problema de la intermitencia es preciso solventar el del almacenamiento que, en el caso de la electricidad de potencia que nos ocupa, puede conseguirse mediante el bombeo de agua en presas de doble vaso, que existen en muy pequeña medida todavía, o bien convirtiendo la electricidad producida en los aerogeneradores en hidrógeno para luego convertir éste de nuevo en electricidad en una pila de combustible cuando se requiera. Justamente, el almacenamiento de energía a partir de fuentes renovables podría ser una de las aplicaciones del hidrógeno como vector energético. Y, desde luego, si se consiguieran nuevos dispositivos de almacenamiento de la electricidad directamente, como la nueva generación de baterías que ya hemos evocado a propósito del transporte, la energía eólica podría contribuir de forma gestionable y todavía más significativa al suministro eléctrico.

Electricidad renovable. El Sol

La radiación solar sobre la superficie terrestre supone, en términos energéticos, del orden de un kW de potencia por metro cuadrado. Promediando sobre todas las horas del año, en un lugar soleado como puede ser el sur de España, equivaldría a la deposición de unos 1.900 kWh por metro cuadrado y por año, lo que equivale al contenido energético de 1,2 barriles de petróleo, o lo que es lo mismo al contenido energético de una capa de 20 cm de petróleo. Lo que, teniendo en cuenta las enormes extensiones de tierra desértica y bien soleada, implica que la energía solar es muy abundante aunque muy difusa. De hecho, en términos de energía primaria, la energía del Sol sobre la superficie terrestre es miles de veces mayor que la energía consumida en todo el planeta.

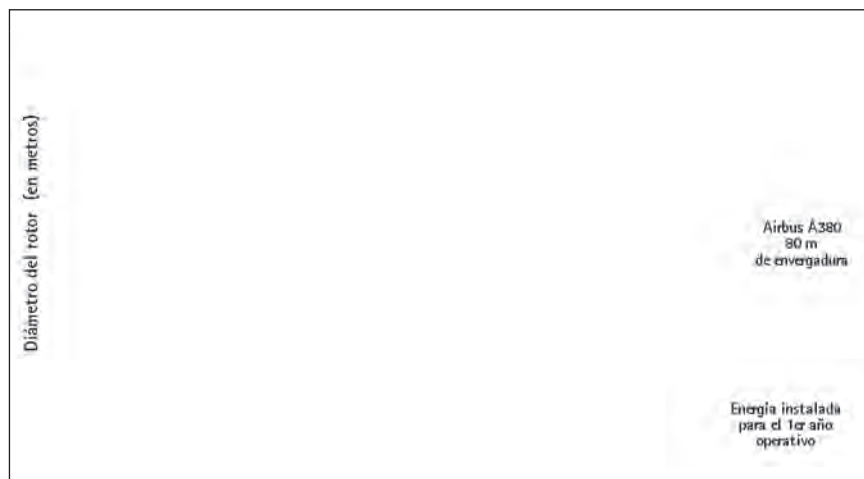


Figura 5. El aumento progresivo de los aerogeneradores. Se detallan potencias en MW y diámetro del rotor (doble de la longitud de las palas) en metros, junto con el primer año en que se puso en operación el primer aerogenerador de cada potencia. Actualmente, existen en Alemania aerogeneradores de hasta 7 MW. Para tener una idea del tamaño de los generadores se comparan con un Airbus 380.

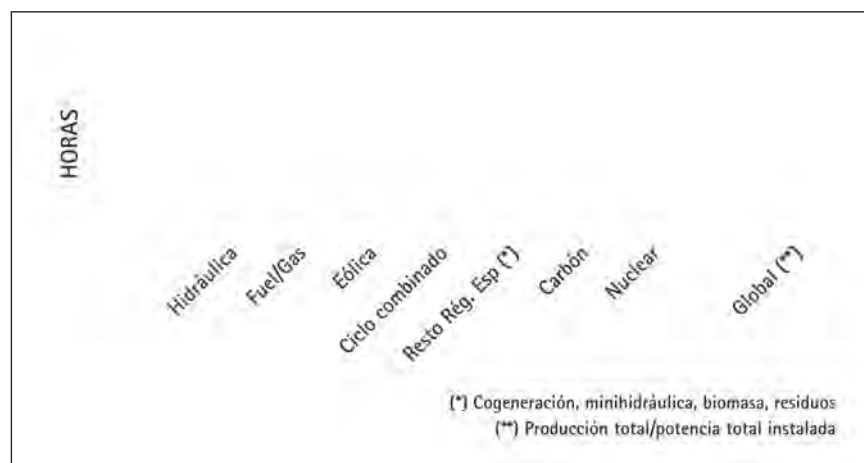


Figura 6. Horas de funcionamiento de las plantas de producción de electricidad por fuente de energía primaria en 2006 (Red Eléctrica Española).

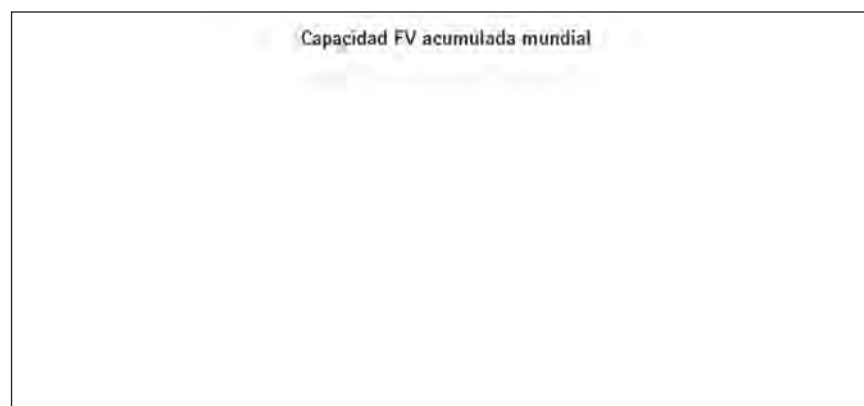


Figura 7. Potencia fotovoltaica instalada en el mundo (EPIA 2008).

Hay dos formas de utilizar la energía del Sol, la primera mediante su conversión directa en electricidad utilizando dispositivos fotovoltaicos hechos de materiales que transforman la energía de los fotones solares en energía de electrones en un conductor, y la segunda mediante la transformación de la energía radiante en calor de alta temperatura y, posteriormente, la conversión de éste en electricidad en una turbina convencional, lo que se conoce como energía solar termoelectrónica.

La energía solar fotovoltaica presenta los mismos inconvenientes del resto de las renovables: precio e intermitencia. El precio se deriva del coste de la fabricación de los dispositivos fotovoltaicos, lo que se traduce en que este tipo de energía es actualmente de los más caros entre las renovables y requiere de un apoyo público considerable. De hecho, en los sistemas basados en tarifas especiales, la energía fotovoltaica es la que recibe mayor subsidio. Por otra parte, la tecnología fotovoltaica es una de las más versátiles y adaptables a los entornos urbanos debido a su carácter distribuido y a no requerir de grandes sistemas de transformación, en contraposición con los dispositivos termoelectrónicos. En cuanto a su difusión, el crecimiento en la

potencia total instalada en el mundo está siendo vertiginoso en los últimos tiempos, como se aprecia en la figura 7.

Alemania es el país con mayor potencia instalada, con 3.800 MW, mientras que en España, con 630 MW a finales de 2007, se ha producido un incremento muy considerable de las instalaciones fotovoltaicas a lo largo de los dos últimos años. Dicho incremento, que no sería sostenible en el tiempo, está asociado a la prima de más de 40 c€/kWh prevista en el Régimen Especial, y al anuncio de que la cuantía de dicha prima se revisaría a la baja en septiembre de 2008. Justamente, la fijación de la prima a la energía fotovoltaica es un buen ejemplo de la importancia de determinar los incentivos de forma inteligente. Si son demasiado bajos en comparación con los costes reales y previsibles, entonces no impulsará el desarrollo de esa tecnología desde el momento en que, como ya hemos observado, una condición necesaria es la creación de un mercado extenso. Pero si es demasiado alta, no cumple el propósito de estimular los avances tecnológicos que puedan ir disminuyendo costes y, en consecuencia, la magnitud de las ayudas públicas asociadas a dichos costes.

Actualmente, la mayor parte de los paneles instalados están compuestos de células fabricadas con oblea de silicio, cristalino o policristalino. El rendimiento medio de estos dispositivos en condiciones de campo, esto es, la fracción de la energía solar depositada sobre la superficie del material que se convierte en electricidad, es del orden del 10 al 15%. Existen otras alternativas para mejorar el rendimiento o para disminuir el coste de las células fotovoltaicas. Una vía es la exploración de otro tipo de materiales y técnicas de deposición, conocidas como sistemas de lámina delgada, que utilizan también silicio, aunque en cantidades menores que en los sistemas convencionales, u otros materiales más exóticos y menos abundantes, que mejoran la conversión fotoeléctrica. O con tecnologías multicapa que permiten la superposición de materiales que son sensibles a diferentes frecuencias del espectro solar con lo que se incrementa el rendimiento total. El objetivo en este apartado es encontrar materiales y procedimientos de fabricación de células que utilicen la menor cantidad posible de materiales, que éstos no sean costosos ni contaminantes y que se integren en distintas aplicaciones, por ejemplo en el ámbito de la edificación, que parecen más adaptadas a este tipo de tecnología. No obstante, se estima que todavía durante muchos años la tecnología dominante será la convencional, basada en oblea de silicio.

Sin embargo, es probable que la mejora en las prestaciones de los sistemas fotovoltaicos tenga lugar a corto plazo debido a las técnicas de concentración que, gracias a dispositivos ópticos, son capaces de hacer incidir sobre una cierta superficie de material fotovoltaico la radiación solar captada sobre una superficie mayor, con lo que aumenta su aprovechamiento. En todo caso, el objetivo fundamental de la tecnología fotovoltaica está en la disminución de costes, que son todavía muy altos en comparación con otras renovables.

Otra forma de utilizar la radiación solar para producir electricidad es mediante la tecnología termoeléctrica. Se trata, en este caso, de concentrar la luz solar sobre un receptor que contiene un fluido que se calienta y posteriormente transfiere ese calor a un sistema de transformación en electricidad en una turbina convencional. Es una tecnología conocida de antiguo, simple en sus principios y robusta, que está experimentando un desarrollo notable en los últimos tiempos, especialmente en España y Estados Unidos. Atendiendo a la forma de los colectores solares y de los receptores, se han diseñado diferentes dispositivos, aunque en lo que sigue vamos a considerar únicamente las dos tecnologías más extendidas, la de colectores cilindro-parabólicos y la de torre o receptor central.

En el primer caso, el calor se concentra sobre un receptor en forma de tubo por el que circula un fluido –normalmente un aceite mineral con propiedades térmicas adecuadas– que se calienta hasta una temperatura de unos 400°C y que, en un intercambiador de calor, genera vapor a alta temperatura y presión que acciona la turbina. Durante la década de los ochenta, tras la segunda gran crisis del petróleo, se construyó en el desierto de Mojave, en California, un conjunto de plantas –el complejo SECS– con una potencia total de 350 MW que han venido funcionando hasta la fecha sin ningún tipo de problema, lo



Figura 8. Planta Acciona-Solargenix de 64 MWe situada en Boulder, Nevada.



Figura 9. Aspecto del campo solar de las plantas SECS situadas en Kramer Junction, California.

que implica que disponemos de una valiosa información acerca del funcionamiento de la tecnología en cuestión. Tras su puesta en marcha, y pasado el momento de crisis, dichas plantas no tuvieron continuación. Por otra parte, de la misma época data la fundación de la Plataforma Solar de Almería (PSA), hoy parte del Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), que es un laboratorio de primera línea mundial en el que se ha venido estudiando el conjunto de las tecnologías termoeléctricas, formando personal y ensayando todos los componentes y dispositivos. La existencia de la PSA es uno de los factores que permite explicar el liderazgo de nuestro país en este campo.

La segunda planta comercial en el mundo que utiliza colectores cilindro-parabólicos está situada en el desierto de Nevada y ha sido construida y está siendo operada por Acciona. Actualmente, hay proyectos para construir este tipo de plantas en España hasta una potencia probable de unos 2.500 MW en los próximos cuatro o cinco años y una cantidad también considerable en Estados Unidos, en general con participación española. Por ejemplo, recientemente se ha adjudicado a Abengoa el llamado proyecto Solana, con dos grandes plantas termosolares por un total de 240 MW ubicadas en Arizona. Las figs. 8 y 9 dan una idea del aspecto de este tipo de plantas y de los espacios requeridos para su instalación.

En nuestro país, entre los numerosos proyectos en curso, está a punto de concluirse Andasol I, la primera planta de 50 MW de un conjunto proyectado por un consorcio en el que participan la empresa Cobra, de ACS, de forma mayoritaria y una empresa alemana, Solar Millenium. La planta de Andasol, situada cerca de Guadix, en Granada, aborda uno de los problemas básicos, ya mencionados, en relación con el uso óptimo de las energías renovables, a saber, el almacenamiento. En este caso, se trata de almacenamiento de calor, que tiene algunas ventajas sobre el de electricidad. En una planta con almacenamiento, una parte del campo solar alimenta, mientras el Sol brilla, el dispositivo de almacenamiento al tiempo que el resto genera el calor que produce electricidad en la turbina. Así, cuando el Sol se pone y la demanda eléctrica sigue alta, es posible seguir generando electricidad a partir de la energía almacenada. En el caso de Andasol I el almacenamiento tiene capacidad para seguir trabajando a la máxima potencia durante 7,5 horas, lo que hace que la planta sea perfectamente gestionable y permite ajustar la entrega de electricidad a la demanda.

El almacenamiento térmico que se está ensayando en este tipo de plantas se basa en grandes cantidades de sales (nitratos) fundidas que almacenan calor elevando su temperatura y lo liberan cuando se enfrían. Es un sistema simple y seguro, aunque requiere movilizar, para las potencias que estamos manejando, cantidades considerables de sales. En particular, en la planta Andasol I se utilizarán 28.500 toneladas de nitratos. Existen otras

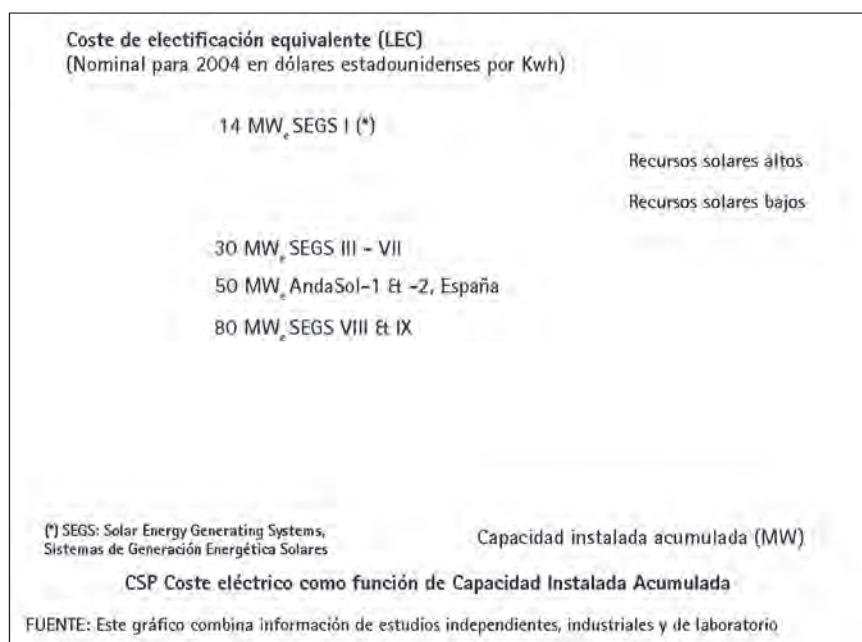


Figura 10. Estimación de la reducción de costes en la electricidad de origen termoeléctrico en función de la capacidad instalada según CSP Global Market Initiative (SolarPaces 2004).

posibilidades de almacenar calor, tales como calor latente en materiales con cambio de fase, en lugar del calor sensible asociado a las diferencias de temperatura, o dispositivos basados en sólidos, alternativas que se irán definiendo y perfeccionando a medida que aumente nuestra experiencia en este campo.

Este tipo de energía de origen solar es también más costoso que la tradicional, aunque menos que la fotovoltaica, siendo la prima establecida en la normativa del Régimen Especial de unos 20 c€/kWh. Y al igual que en el resto de las renovables, es de esperar una reducción de costes a medida que el mercado se expanda. De acuerdo con los estudios de SolarPaces, los costes convergerán con los de las energías convencionales cuando se hayan instalado del orden de 15.000 MW, como puede verse en la figura 10.

Para que esto ocurra son precisos algunos avances tecnológicos, especialmente en la manufactura de los tubos absorbentes, al tiempo que se diversifica el mercado de proveedores, hoy tan estrecho que es difícil que se desarrollen los mecanismos de competencia comercial imprescindibles para la mejora de los procesos de fabricación. Otro de los desarrollos esperados se refiere al fluido caloportador. Como ya se ha dicho, actualmente se utiliza un aceite térmico mineral que tiene el inconveniente de que más allá de una cierta temperatura –del orden de 450°C– se descompone, por lo que no es posible aumentar la temperatura de trabajo y, por tanto, el rendimiento de la conversión de calor en electricidad. Por otra parte, se trata de una sustancia engorrosa de manejar y contaminante. En este sentido, existen programas de investigación avanzados cuyo propósito es la sustitución del aceite por otro fluido, como agua o un gas, que permita aumentar

la temperatura de trabajo y simplifique el diseño de las plantas, abaratando su coste. En estos programas están involucrados grupos de investigación, alemanes y españoles, que trabajan en la PSA, así como las más importantes empresas del sector (véase, por ejemplo, Zarza 2008). En resumen, los retos que plantea la utilización de estas tecnologías se centran en la optimización de tubos, fluido caloportador, sistemas de almacenamiento y colectores, así como en la ampliación del mercado global sobre la base de incentivos públicos.

Otra tecnología en desarrollo dentro del capítulo de la energía solar termoeléctrica es la que se basa en un receptor central situado en la parte superior de una torre. Un campo de helióstatos rectangulares concentra la radiación solar sobre el receptor en cuestión, del que se extrae el calor generado mediante algún fluido líquido o gaseoso. Las primeras plantas comerciales en operación han sido construidas en Sanlúcar la Mayor (Sevilla) por Abengoa, las PS-10 y PS-20, de 11 MW y 20 MW de potencia, respectivamente. Por el momento, los costes son superiores a los de la generación basada en colectores cilindro-parabólicos y están en un grado de desarrollo algo más atrasado. Pero presentan algunas ventajas como poder operar a temperaturas mayores y adaptarse a terrenos más accidentados. El proceso de mejora y optimización que se está iniciando es similar al descrito anteriormente, incluyendo los dispositivos de almacenamiento térmico, conceptualmente similares.

La fisión nuclear

Junto a los combustibles fósiles y las renovables, la energía de fisión nuclear es una fuente de energía esencial hoy en los países más desarrollados. En Europa el 30% de la electricidad producida es nuclear, siendo este porcentaje del 20% en España. La energía nuclear tiene algunas ventajas que la hacen atractiva como parte del menú energético del futuro. Las principales son la independencia de cualquier condición climática o ambiental, que permite que una planta opere un porcentaje muy elevado de las horas del año, como puede verse en la figura 6. Este hecho explica que el sector nuclear en España, con 7.700 MW instalados, haya generado en 2007 aproximadamente el doble de la electricidad de origen eólico que cuenta con una potencia total de 15.100 MW. Otro factor positivo a tener en cuenta es la relativa independencia de las oscilaciones del precio del uranio, ya que éste, a lo largo de la vida útil de la planta, supone apenas un 6% de los costes totales de construcción y operación. En la figura 11 se compara el coste de la materia prima en las plantas nucleares con las de otras fuentes de energía convencionales.

Por otra parte, se trata de un sector industrial con una considerable experiencia en seguridad, a pesar de la existencia de una opinión generalizada en contra. De hecho, los protocolos de seguridad más avanzados y exigentes han surgido precisamente de la industria nuclear.

Sus inconvenientes son conocidos. Desde el punto de vista económico, las enormes inversiones necesarias en la fase de construcción, con un dilatado periodo de amortización, que es la contrapartida del reducido coste del combustible. Desde el punto de vista ambiental o de seguridad, la gravedad potencial de los accidentes en el funcionamiento de las plantas, aunque éstos sean escasos y, sobre todo, la generación de residuos radiactivos difíciles de gestionar y custodiar. El problema de los residuos es, seguramente, el inconveniente más serio y, en la opinión pública, sin duda ha predominado sobre los aspectos más positivos de esta tecnología energética por lo que merece una consideración especial.

Los residuos son, a grandes rasgos, de dos clases. De corta duración y de larga duración. Los primeros tienen típicamente una vida media de unos 30 años –la vida media o semiperiodo se define como el tiempo que debe transcurrir para que la actividad de un material disminuya a la mitad–. Para este tipo de residuos, que son los que se generan en mayor cantidad, la solución aceptada universalmente es el almacenamiento en un depósito durante el tiempo necesario para que la actividad haya disminuido hasta el nivel de la radiactividad natural de fondo. El Cabril, en Córdoba, es un ejemplo típico de este tipo de almacenamiento y, gestionado adecuadamente, sus efectos sobre el medio ambiente son inapreciables.

El problema más grave son los residuos de vida media muy larga, decenas o cientos de miles de años, que se concentran sobre todo en las barras de combustible usado. Algunos países han optado por construir Almacenes Geológicos Profundos (AGP), en condiciones de estanqueidad suficientes como para garantizar la estabilidad de los residuos depositados en periodos geológicos. Está clara la dificultad no sólo de encontrar lugares que cumplan con las condiciones físicas requeridas, sino también de aceptación por parte de la opinión pública. Otros países, como España, se orientan a construir un Almacén Temporal Centralizado (ATC) en superficie, que permita la custodia de los residuos en condiciones de seguridad durante periodos mucho más cortos, del orden de un siglo, mientras se ponen a punto técnicas de eliminación de los residuos, o su transformación en materiales inertes. Y es que la gestión o eliminación de los residuos es uno de los problemas cuya resolución es más acuciante si queremos que la energía nuclear tenga un futuro. Se conoce el principio de dicha transformación, lo que se conoce con el nombre de técnicas de Separación y Transmutación, pero su desarrollo es todavía muy incipiente. Y lo es por su complejidad tecnológica, pero también por la dificultad de experimentar en tecnologías nucleares en un entorno de fuerte oposición pública.

En realidad, los desarrollos asociados a la neutralización de los residuos más peligrosos están muy ligados a lo que se conoce como reactores de cuarta generación. Existen en este momento 439 reactores comerciales en funcionamien-

to en el mundo, de los que 104 están en Estados Unidos y 59 en Francia, con una potencia instalada de 373.000 MW, y 38 más en construcción en Finlandia, Francia, los países del este de Europa y Asia (World Nuclear Association 2008). Todos ellos son de segunda o tercera generación, operando con neutrones térmicos –lentos– y utilizando como combustible el isótopo ^{235}U , que es muy poco abundante en la naturaleza, sólo el 0,7% del uranio natural. Las líneas más prometedoras de la cuarta generación operan con neutrones rápidos y pueden utilizar como combustible la mayoría de los residuos existentes así como el ^{238}U , que es el isótopo de uranio más abundante –el otro 99,3%– e incluso el torio, que es todavía más abundante, en una alternativa explorada de forma prominente en la India. Los reactores de cuarta generación y los dispositivos que incorporan las tecnologías de neutrones rápidos, por ejemplo los Sistemas Asistidos con un Acelerador –ADP, de sus siglas en inglés–, podrían resolver potencialmente muchos de los problemas asociados con los residuos. Y serían inmunes a una eventual escasez de combustible convencional a largo plazo –si utilizáramos todo el uranio y no sólo su isótopo fisible y poco abundante, las reservas se multiplicarían automáticamente por un factor de más de cien.

Los retos indiscutibles en el sector nuclear son, por lo tanto, el tratamiento de los residuos y la cuarta generación, ambos aspectos relacionados desde el punto de vista de la tecnología. Pero los avances en este terreno llevan su tiempo y no estarán disponibles, en condiciones de explotación comercial, antes de veinte o treinta años, con lo que el sector nuclear, en la mayoría de los países occidentales con las señaladas excepciones de Francia y Finlandia, se encuentra con la dificultad de un improbable resurgimiento durante todo este tiempo, lo que puede incidir en una pérdida de conocimientos y capacidades técnicas. En contraste, en muchos otros países del mundo, en particular en Asia, se seguirán construyendo y operando reactores nucleares de segunda y tercera generación.

Conclusiones

Ante la situación descrita en los párrafos precedentes, no parece realista ni aconsejable prescindir de ninguna de las fuentes de energía disponibles, con las debidas precauciones y en los tiempos en los que la tecnología lo permita. A corto plazo urge preparar sustitutos de los derivados del petróleo para el sector del transporte, entre los que no podemos dejar de considerar los biocombustibles de segunda generación. Respecto del carbón, que seguirá siendo una fuente abundante aunque potencialmente muy contaminante, se impone avanzar hacia su utilización con captura y secuestro de CO_2 .

Pero quizá el reto más importante en estos momentos sea impulsar las energías renovables de forma que lleguen a suponer una fracción significativa del total, situación de la que estamos muy lejos hoy y en la que España ocupa un papel de vanguardia. El viento ha demostrado sus poten-

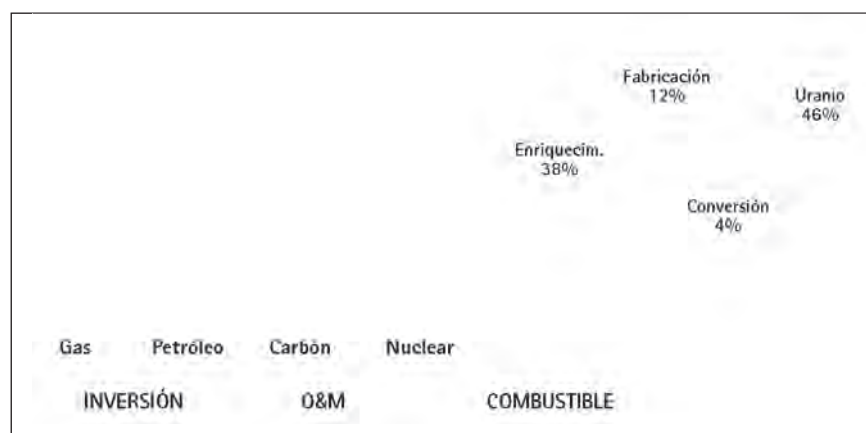


Figura 11. Distribución de costes en distintos tipos de plantas de producción de electricidad.

cialidades como fuente masiva de energía y debe seguir ampliando su presencia en el mercado global. El Sol, más abundante aunque con los problemas de dispersión ya evocados en el texto del artículo, deberá ir ocupando, en algún momento del futuro próximo el papel de energía renovable dominante, realmente masiva, sostenible e ilimitada. Para ello habrán de resolverse los problemas tecnológicos que limitan su difusión y afectan al elevado precio que hoy tiene, y se necesitará un decidido apoyo público. Por razones de gestionabilidad de las energías renovables y también atendiendo al futuro del sector del transporte, las tecnologías de almacenamiento de energía ocupan ya un lugar destacado en los programas de investigación energética,

hasta el punto de que no es concebible un esquema sostenible sin un dominio suficiente de este tipo de tecnologías.

Desafortunadamente, la fusión nuclear se incorporará más tarde y no es verosímil que contribuya a aliviar la situación en las próximas décadas. Pero los reactores de fisión existen, han sido probados y han evolucionado hacia diseños cada vez más seguros y con un mejor aprovechamiento del combustible. No creo que sea razonable, en una situación de crisis energética, prescindir de esta fuente de energía, aunque su supervivencia depende en gran medida de la percepción pública. A corto plazo el problema se plantea en términos de prolongación de la vida útil de los reactores existentes y de su reemplazamiento por otros de generación III; sin embargo, el desafío fundamental en este apartado es el avance hacia los reactores rápidos de generación IV que permitirán reciclar los residuos y utilizarán el combustible de forma óptima.

No hay soluciones milagrosas que puedan resolver de un plumazo el problema del aprovisionamiento energético de la humanidad. Necesariamente requerirá abordarlo desde todas las perspectivas posibles y actuar desde el punto de vista tecnológico, pero también político y financiero, en todas y cada una de las fuentes disponibles. Sin olvidar el aspecto educativo y de divulgación, tan necesario en una situación en la que la mayoría de la población da por resuelto el problema energético y por descontada la seguridad en el suministro, pero no acepta las servidumbres que inevitablemente conlleva la producción de energía, tanto de tipo económico como de afectación de territorio.

Bibliografía

- Ballesteros, M. «Estado del desarrollo tecnológico de los biocombustibles». *Energía*, vol. 202 (2007): 24-28.
- British Petroleum. *BP Statistical Review of World Energy* (junio 2008).
- CIEMAT. *Análisis del ciclo de vida de combustibles alternativos para el transporte*. 2005.
- Club Español de la Energía. *Balance energético de 2007 y perspectivas para 2008*. Madrid, 2008.
- , *Energía: Las tecnologías del futuro*. 2008a.
- European Commission. *The Sustainable Nuclear Energy Technology Platform*. 2007.
- European Photovoltaics Industry Association. *Solar Generation V* (septiembre 2008).
- European Wind Energy Association. *www.ewea.org*. 2008.
- Fundación para Estudios sobre la Energía. *El futuro del carbón en la política energética española*. Madrid, 2008.
- Global Wind Energy Council. *Global Wind 2007 Report*. 2008.
- International Energy Agency. *World Energy Outlook*. 2006.
- , *Key World Energy Statistics*. 2008.
- Signes, V. et al. *Procedimiento para la revalorización energética de la fracción orgánica de residuos orgánicos e instalación*. Solicitud de patente internacional ES2008/000077, 2008.
- Socolow, R. H. «Can we bury global warming?». *Scientific American* (julio 2005): 39-45.
- Solar Paces. *The CSP Global Market Initiative*. 2004.
- United Nations. *Human Development Report*. 2006.
- World Nuclear Association. *www.world-nuclear.org*. (septiembre 2008).
- Worldwatch Institute. *Biofuels for transport*. Earthscan, 2007.
- Zah, R. et al. *Ökobilanz von Energieprodukten: Ökologische Bewertung von Biotreibstoffen*. Switzerland: Empa, St. Gallen, 2007.
- Zarza, E. et al. *Almería GDV: The first solar power plant with direct steam generation*, 2008.

cambio de clima en el planeta tierra

SERGIO ALONSO

*A mi esposa y compañera Mercedes, in memoriam
A nuestros hijos Aurora, Patricia y Carlos*

Introducción

Este año 2008 se celebra el vigésimo aniversario del establecimiento del Panel Intergubernamental del Cambio Climático –IPCC, siglas correspondientes a su denominación en inglés: Intergovernmental Panel on Climate Change–. Su creación se debió a un acuerdo entre la Organización Meteorológica Mundial –órgano de Naciones Unidas– y el Programa de las Naciones Unidas para el Medio Ambiente y tuvo como objetivo proporcionar información científica independiente sobre la problemática del Cambio Climático, en principio, dirigida a los políticos. Casi diez años antes, en la primera Conferencia Mundial sobre el Clima, se había llamado la atención sobre el incremento de las actividades humanas, indicando que podrían producir alteraciones climáticas a escala regional e incluso planetaria. Unos años más tarde se evaluó el papel del CO₂, y otros gases capaces de contribuir al denominado efecto invernadero, en variaciones del clima. También se proclamó la necesidad de disponer de juicios científicos objetivos, equilibrados y coordinados internacionalmente, capaces de aportar luz sobre las consecuencias que un aumento de la concentración en la atmósfera terrestre de los gases con efecto invernadero podría tener en el clima del planeta y el impacto socio-económico que se podría derivar. Esta preocupación ambiental que oficialmente se hizo pública hace unos treinta años, aunque en realidad es anterior, condujo al establecimiento del IPCC en 1988. En 2007, el Comité Nobel decidió que el premio Nobel de la Paz debería ser

compartido, en dos partes iguales, entre el Panel Intergubernamental del Cambio Climático (IPCC) y Albert Arnold (Al) Gore Jr. «por sus esfuerzos en aumentar y difundir el conocimiento sobre el cambio de clima inducido por el hombre, y por poner la base para las medidas que son necesarias para contrarrestar dicho cambio».¹

Algunos de los términos que han surgido en esta introducción serán tratados con algo de detalle más adelante, pero conviene que otros queden precisados al máximo desde un principio. En primer lugar hay que indicar que el planeta ha experimentado cambios de clima de forma continuada. Se tiene la seguridad de que el clima del pasado ha sido diferente del presente y que en el futuro habrá también un clima distinto al actual. La terminología, en un principio, resultaba algo confusa al coexistir denominaciones como variación del clima, variabilidad del clima, cambio de clima y cambio climático. Final y desafortunadamente, se han mantenido dos acepciones. En términos científicos, cambio de clima, o cambio climático, representa cualquier cambio que experimente el clima del planeta independientemente de su causa. Esta opción es la utilizada, por ejemplo, por el IPCC. Sin embargo, la Convención Marco de Naciones Unidas sobre el Cambio Climático, surgida de la denominada Cumbre de Río de 1992, y el Protocolo de Kioto –establecido a partir de la Convención–, utilizan la misma terminología para referirse al cambio del clima atribuido directa o indirectamente a la actividad humana y que se superpone a la variabilidad natural. Esto hace que, cuando se hable

¹
Párrafo tomado del anuncio oficial de concesión del premio por parte de la Fundación Nobel.

de Cambio Climático, se deba tener suficiente cuidado para que quede claro cuál de las dos acepciones se está empleando. Obsérvese, por ejemplo, que en la declaración de la Fundación Nobel se precisa que el cambio de clima al que se refiere es el inducido por el hombre. Veremos más adelante que este cambio de clima, de origen antrópico, se explica en términos de la intensificación del efecto invernadero natural. Dicha intensificación deriva del cambio de composición que ha experimentado nuestra atmósfera como consecuencia de la actividad humana.

El contenido de esta contribución incluye, en el próximo apartado, las razones por las que cambia el clima del planeta, bien sean naturales o antrópicas. En el apartado 3 se pasará revista a las observaciones recientes de los cambios experimentados por el clima, dejando para el siguiente la argumentación, basada en la simulación numérica del clima, de la atribución de dichos cambios a la actividad humana. En el apartado 5 se dan algunas indicaciones sobre el uso de modelos ejecutados por ordenador para simular el clima de la Tierra. En base a la confianza que se otorga a los modelos para reproducir el clima computacionalmente, en el apartado 6 se aborda la obtención de escenarios climáticos para el futuro. En el apartado 7 se presentan las conclusiones y finalmente la bibliografía empleada.

¿Por qué cambia el clima?

El clima es algo dinámico, cambiante, incluso irreplicable, consecuencia de la energía que recibe la Tierra del Sol y de los intercambios de energía entre partes diferentes de lo que se llama Sistema Climático, que podemos entender como sinónimo de planeta Tierra. Esas partes o subsistemas son:

- a) La atmósfera, envoltura gaseosa del planeta, allí donde percibimos el clima.
- b) La hidrosfera, formada por océanos, mares, lagos, etc.
- c) La litosfera, corteza sólida emergente de los continentes, allí donde vivimos.
- d) La biosfera, formada por todos los seres vivos, incluido el hombre.
- e) La criosfera, formada por los hielos que cubren parte de océanos y continentes.

Desde un punto de vista amplio, el clima se define como el estado del Sistema Climático, incluyendo sus propiedades estadísticas. Es precisamente esto lo que relaciona esta definición de clima con la más clásica, y restringida, consistente en considerarlo como una descripción estadística de variables ambientales (por ejemplo, temperatura, viento, humedad en superficie o precipitación) usando valores medios y medidas de dispersión en intervalos de tiempo largos, muy superiores a los típicos del tiempo atmosférico.

Los subsistemas indicados del Sistema Climático tienen dinámicas muy diferentes. Mientras unos sufren cambios continuos apreciables (la atmósfera, por ejemplo, con su sucesión de tiempos –soleados, nubosos, ventosos, lluviosos, etc.– tan diferentes) otros lo hacen muy lentamente;

tan lentamente, en algunos casos, que, para la vida del hombre o de varias generaciones humanas, no tiene sentido considerar su variabilidad (como sería el caso de la litosfera, a excepción de la capa de suelo más superficial). Cuando la energía que recibimos del Sol alcanza la Tierra se distribuye entre todos los subsistemas y se intercambia entre unos y otros, atendiendo cada uno de ellos a su propia dinámica. De la diferencia entre estos intercambios surge la gran variedad de climas de las distintas regiones de nuestro planeta, que tan bien conocemos y que son la manifestación de la variabilidad espacial del clima.

Pero el clima también presenta variabilidad temporal. La energía que proviene del Sol no llega en igual cantidad en momentos diferentes, ni los subsistemas del Sistema Climático se comportan siempre exactamente igual. En consecuencia, no hay que esperar que los flujos de energía que se establecen coincidan invariablemente en el transcurso del tiempo. En ciertos intervalos de tiempo sus estadísticas pueden, más o menos, coincidir, pero no tiene sentido pensar que eso tenga que ser siempre así.

A continuación vamos a analizar con algo de detalle el origen de la variabilidad o, dicho de otra forma, las causas de que cambie el clima de la Tierra. De ellas unas son naturales y otras no, o sea, tienen que ver con la actividad humana. El nivel de conocimiento que se tiene de los mecanismos que veremos a continuación es, en general, elevado, pero no hay que perder de vista que siempre que hay déficit de conocimiento –y de hecho siempre lo hay– se tiene un cierto grado de ignorancia, lo que da lugar a una incertidumbre en la interpretación de los fenómenos observados.

En primer lugar hay que empezar hablando del Sol y su relación con la Tierra. Todos los procesos que se producen en el Sistema Climático, en el planeta, tienen su fuente casi exclusiva de energía en el Sol. Esa energía viaja por el espacio como radiación –denominada solar o de onda corta– y alcanza la Tierra, que la intercepta, sea cual sea la posición en que se encuentra en su órbita en las diferentes épocas del año. No toda la energía interceptada es aprovechada por el Sistema Climático pues una fracción de ella –denominada albedo– es devuelta al espacio por diferentes procesos de reflexión en las nubes y en la superficie de la Tierra principalmente. El albedo planetario resulta ser del orden de un 30%. Por último, la radiación que no es absorbida en la atmósfera alcanza la superficie, con lo que ésta se calienta y a su vez emite radiación (en este caso denominada radiación terrestre o de onda larga). Una buena parte de esa radiación es absorbida por la atmósfera que la reemite hacia la superficie o hacia arriba, devolviendo así energía al espacio. En conjunto, para todo el planeta y por término medio en el tiempo existe balance de energía, pero no en sus diferentes partes ni en todo momento. Son justamente esos desajustes los que condicionan el clima. Para mayor información se puede consultar Kiehl y Trenberth (1997).

¿Cómo se puede alterar el balance de energía? Según lo dicho habría tres razones:

- a) Cambios en la energía interceptada por la Tierra. Éstos pueden ser debidos a cambios en la emisión de radiación por el Sol como consecuencia de la propia actividad solar y a cambios en la posición de la Tierra en su órbita alrededor del Sol.
- b) Cambios en el albedo terrestre. Se deberían ahora a cambios en la nubosidad –tanto en cobertura como en tipología–, cambios en las propiedades reflectoras del suelo –tipos de suelos o vegetación– y cambios en la materia particulada que se encuentra en suspensión en la atmósfera, conocida con el nombre de «aerosoles».
- c) Cambios en el flujo de energía de onda larga de la Tierra al espacio. En este caso los cambios serían debidos a la modificación de las propiedades absorbentes de la atmósfera como consecuencia de cambios en la composición atmosférica.

Los cambios en la actividad solar están referenciados. Posiblemente el más popular sea el denominado Mínimo de Maunder, que se estima ocurrió entre 1350 y 1850, coincidente con la denominada Pequeña Edad de Hielo (Hoyt, Schatten y Nesmes-Ribes 1994; Eddy 1976). Desde esa época se estima que la radiación puede haber aumentado entre un 0,04 y un 0,08%, con un incremento del 0,05% desde 1750 a la actualidad (Wang, Lean y Sheeley 2005).

Pero la Tierra no ocupa una posición fija en relación al Sol; describe muy aproximadamente una trayectoria elíptica, con el Sol en un foco, cuya excentricidad cambia en el transcurso del tiempo con una periodicidad de unos cien mil años. Esto hace que la Tierra se encuentre a una distancia del Sol que no va siendo igual año tras año al recorrer su órbita, que además es cambiante. Por otra parte, la inclinación del eje del mundo con respecto al plano de la órbita –oblicuidad– no es constante, sino que, como si la Tierra fuera una gran peonza, la prolongación de su eje de rotación señala puntos diferentes de la cúpula celeste en ciclos de alrededor de cuarenta y un mil años. Además, la elipse orbital cambia de orientación en el espacio, dando lugar a lo que se llama precesión de los equinoccios. Esto hace que las estaciones astronómicas se den en diferentes lugares de la órbita con periodicidades aproximadas de diecinueve mil y veintitrés mil años. El resultado final es que, aunque fuera constante la energía emitida por el Sol, es diferente la energía incidente en el sistema y, además, se distribuye de forma diferente sobre la superficie del planeta. Lo anterior constituye la llamada teoría de los ciclos de Milankovitch, la cual permite explicar, junto con algún mecanismo interno, la sucesión de las eras geológicas (Berger 1988).

Los procesos descritos son externos al Sistema Climático y no dependen en absoluto de la actividad humana. Otra posible causa de cambio de clima en el planeta, también externo y natural, pero que no guarda ninguna relación con la radiación solar recibida en la Tierra, es el impacto de meteoritos o cometas. Se trata de algo difícilmente predecible, pero de consecuencias importantes si el tamaño del bólido es suficientemente grande. Su impacto

contra la superficie del planeta puede originar una nube de polvo y/o de agua de tal magnitud que la radiación solar incidente no alcance el suelo con la intensidad anterior al impacto. En esas condiciones, la temperatura puede descender de una forma apreciable, dando lugar a un cambio en el clima. La extinción de algunas especies, entre ellas los dinosaurios, en el llamado límite K/T, parece que tuvo este origen (Álvarez et al. 1981).

Esa causa, que podemos calificar de excepcional, nos sirve para introducir aquellas relacionadas con el albedo. Tras el impacto se tuvo que producir un aumento considerable del albedo al aumentar la cantidad de aerosoles –materia particulada– en la atmósfera, haciendo que una fracción muy elevada de la radiación solar incidente volviera al espacio. En consecuencia el Sistema Climático dispondría repentinamente de mucha menos energía para calentar el suelo y, por lo tanto, se alteraría el anterior balance de radiación. La consecuencia tuvo que ser una disminución de la temperatura al nivel del suelo. Sin llegar a esos extremos, cada vez que se produce una erupción volcánica ocurre algo similar. Su efecto sobre la temperatura se ha observado tras las grandes erupciones; depende de la intensidad de la erupción y de la altura que alcanzan en la atmósfera las partículas generadas, pudiendo durar varios años. Este efecto ha sido ampliamente estudiado (véase, p.e., Yang y Schlesinger 2002).

Los aerosoles considerados hasta el momento son de origen natural, pero en la atmósfera terrestre se encuentran, junto a los naturales, otros muchos que han sido originados por la actividad humana. En general, hacen disminuir la calidad del aire y, además, muchos de ellos dan lugar a problemas de salud. Desde un punto de vista climático tienen dos efectos. Uno directo sobre el albedo, que da como consecuencia una disminución de la temperatura, y otro, indirecto, al modificar las condiciones en que se forman las nubes y su duración. El resultado final de este efecto indirecto no es bien conocido y constituye, hoy por hoy, una causa de incertidumbre.

El papel de las nubes en cuanto al albedo depende de la cobertura nubosa, del tipo de nube y de su tiempo de vida. Así, las nubes altas como, por ejemplo, los cirrostratos, dejan pasar la radiación solar pero absorben la terrestre, mientras que las nubes medias, por ejemplo los altocúmulos, impiden casi completamente el paso de la radiación solar. En el primer caso el efecto sobre la temperatura sería un incremento, mientras que en el segundo se tendría un enfriamiento.

El albedo también depende, como se ha dicho, de las propiedades reflectoras de la superficie del planeta. No es lo mismo considerar una superficie helada –alto albedo, del 70 al 90%–, que un suelo desnudo, una pradera, o la superficie marina –bajo albedo, < 10%–. La diferente tipología de terreno y de uso del suelo hace que el tratamiento climático de la superficie sea un problema complejo y fuente de incertidumbre.

Llegado este punto, no se puede dejar de comentar un tipo de comportamiento característico del Sistema Climático. Muchas veces actúan los efectos de un proceso sobre las propias causas, estableciéndose una especie de comportamiento cíclico, sin fin, que recibe el nombre de realimentación. Las realimentaciones son típicas en los llamados sistemas no lineales, o dinámicos, y el Sistema Climático lo es. Este ejemplo es de los sencillos: supongamos que, por cualquier causa, la temperatura superficial del planeta aumenta. Una de las consecuencias sería la fusión parcial de los hielos. El albedo de la superficie disminuiría, con lo que se produciría menor reflexión de la radiación solar, habría más energía disponible para el sistema y la temperatura se incrementaría. El calentamiento adicional daría lugar a mayor fusión de los hielos, reduciéndose más el albedo, y así progresivamente. En este caso se tiene una realimentación, positiva, que se conoce con el nombre de realimentación hielo-albedo y que fue identificada ya en el siglo XIX (Croll, 1890). En el Sistema Climático hay otras muchas positivas, como ésta, pero también las hay negativas. Cuando estos procesos de realimentación actúan simultáneamente resulta muy difícil saber con detalle el resultado, aunque es evidente que existe. La única posibilidad de tratar el problema es mediante la simulación numérica de dichos procesos.

La última forma de modificar el balance de radiación que queda por comentar, bien podría haber sido la primera; se trata de la principal para explicar el cambio de clima que está experimentando el planeta en nuestros días.

En primer lugar se considerará el papel que juega la atmósfera en los intercambios de radiación solar y terrestre, lo que se conoce con el nombre de Efecto Invernadero (EI). Ya se ha comentado que parte de la radiación que proviene del Sol, aproximadamente un 30%, es reflejada hacia el espacio. Si la Tierra no dispusiera de atmósfera, la superficie del planeta se encontraría a una temperatura media de -18°C , justo la necesaria para mantener el equilibrio energético resultante del balance entre la radiación solar que penetra y la radiación terrestre –infrarroja– que la Tierra emitiría a esa temperatura. La Luna, que no posee atmósfera, se encuentra a una temperatura media como la indicada. Sin embargo, en la Tierra, al existir atmósfera, las cosas son radicalmente diferentes. Los constituyentes atmosféricos absorben relativamente poca radiación solar –sobre todo en ausencia de nubes– pero algunos son muy buenos absorbentes para la radiación infrarroja que emite la Tierra y la propia atmósfera. En consecuencia, se produce un calentamiento en las capas bajas de la atmósfera, que modifica el balance de radiación, alcanzando una temperatura media de 15°C al nivel de la superficie. Este comportamiento de la atmósfera, radiativamente diferente para la radiación solar que para la terrestre, es el EI, nombre recibido por guardar cierta semejanza con el comportamiento de esa estructura. El principal responsable del EI es el vapor de agua –aproximadamente en un

80% del efecto total– y el segundo, a bastante distancia, el dióxido de carbono (CO_2). El EI –al que muchas veces se le añade el calificativo «natural»– es decisivo en el clima que posee el planeta y ha permitido la vida, al menos en la forma que la conocemos. A los gases que contribuyen al EI se les llama gases con efecto invernadero (GEI). Tras lo dicho, debe resultar evidente que al EI contribuyen también los aerosoles y que el papel de las nubes se puede discutir también en estos términos.

Cualquier cambio que se produzca en la composición atmosférica, o en la concentración de sus componentes, altera las propiedades de absorción y, en consecuencia, el EI. La composición de la atmósfera, desde que la Tierra es Tierra, ha sido cambiante. Ahora predominan nitrógeno (N_2) y oxígeno (O_2), aunque los mayores contribuyentes al EI son el vapor de agua –cuya concentración no supera el 4% en volumen de la atmósfera– y el CO_2 –con una concentración mucho menor, en la actualidad del orden de unas 385 ppm^2 –. Si la composición atmosférica cambia, se modifica el EI y, en consecuencia, la temperatura media superficial del planeta. Anteriormente a la revolución industrial, la concentración media global del dióxido de carbono era del orden de 280 ppm mientras que en la actualidad es, como se ha dicho, de unas 385 ppm. En estas condiciones, el EI natural que se da en el planeta está siendo modificado desde la Revolución Industrial. Como la concentración de CO_2 ha aumentado, pero también la de otros GEI, como metano, óxido nitroso, CFCs, etc., el EI se ha intensificado, ha habido más energía disponible en las capas bajas de la atmósfera y, por tanto, se han venido dando las condiciones para que se produzca un calentamiento a escala planetaria. Esto no se trata de una especulación moderna; a finales del siglo XIX el nobel Svante Arrhenius llegó a hacer una estimación del efecto sobre la temperatura de aumentar o disminuir un 40% la concentración de CO_2 indicando que se podrían producir retrocesos o avances en los glaciares (Arrhenius 1896). En realidad, ya desde final del siglo XVII se conocía el diferente comportamiento de ciertas sustancias frente a la radiación solar y a la radiación terrestre, lo que constituye la base del EI.

Analizando el aire de las burbujas atrapadas en los testigos extraídos del hielo polar se puede obtener información sobre la evolución de la concentración de los GEI en épocas pasadas y compararla con la actual. En la fig. 1 están representados los valores de las concentraciones de dióxido de carbono, óxido nitroso y metano para los últimos 650.000 años. Se observa que los valores actuales exceden por mucho a los previos, incluso en los periodos cálidos de las glaciaciones, marcados en bandas sombreadas en la fig. 1. En la parte baja de esa figura se representa también la variación de deuterio, δD , en el hielo ártico, que se utiliza como medida indirecta de las variaciones de temperatura. Nótese los valores δD de los anteriores periodos cálidos y del actual y la gran diferencia en las concentraciones de GEI. A diferencia del presente, en que la relación GEI-tem-

2 Partes por millón, medida de concentración para constituyentes poco abundantes. Equivale a una fracción molar de $\mu\text{mol/mol}$. De forma semejante, una fracción molar de nmol/mol se representa por ppb (partes por «billion» –mil millones–) y pmol/mol por ppt (partes por «trillion» –billón en castellano–). Si se toma en consideración el comportamiento no ideal de los gases, a veces se utilizan concentraciones en volumen (ppmv, ppbv, pptv), diferentes de las anteriores.



Figura 1. Variaciones de deuterio (D) en el hielo ártico, representativo de la temperatura local, y las concentraciones atmosféricas de los gases de efecto invernadero, dióxido de carbono (CO_2), metano (CH_4), y óxido nitroso (N_2O) en el aire atrapadas en los testigos de hielo y de mediciones atmosféricas actuales. Los datos cubren 650.000 años y las líneas sombreadas representan períodos interglaciares cálidos, el actual y los pasados (Fuente: IPCC 2007).

temperatura está claramente establecida y el origen antrópico del cambio en los GEI probado, en el pasado hay muchos aspectos de esa relación que aún se están investigando. Se piensa que en el Cuaternario los cambios en la concentración de CO_2 se pudieron producir por la actuación simultánea de procesos biológicos y químicos en el océano y también conducidos por cambios de temperatura (Köhler et al. 2005). En épocas anteriores, hace millones de años, en que sí se dieron situaciones con concentraciones de CO_2 muy superiores a las actuales se piensa que fueron procesos tectónicos, como la actividad volcánica, los que determinaron los cambios de concentración (Ruddiman 1997).

Recientemente, como resultado del proyecto de investigación europeo EPICA, se ha ampliado el rango de tiempo hasta 800.000 años. Se mantienen las mismas conclusiones en relación a las concentraciones de GEI que se han indicado en la descripción de la fig. 1 para los últimos 650.000 años (Lüthi et al. 2008; Loulergue et al. 2008).

En la fig. 2 están representadas las variaciones de concentración de CO_2 , CH_4 y N_2O pero para un alcance temporal menor (paneles a, b y c). En la escala de la izquierda de esos paneles se da la concentración del correspondiente GEI, mientras que en la escala de la derecha se

representa el denominado forzamiento radiativo, equivalente de la intensificación del EI que implica el aumento de concentración del GEI expresado en unidades de radiación (Wm^{-2}). De esos tres paneles se deduce que el cambio experimentado por los GEI tras la Revolución Industrial no tiene precedente reciente: mientras que la concentración atmosférica de CO_2 aumentó sólo 20 ppm durante los 8.000 años previos a la industrialización, a partir de 1750 la concentración de CO_2 ha aumentado más de 100 ppm. Aproximadamente las dos terceras partes de ese incremento se deben a la quema de combustibles fósiles y el tercio restante al cambio de usos de la tierra. En el panel d se representa el ritmo de cambio del forzamiento combinado de los mismos tres GEI, lo cual da como valor integrado $1,66 \text{ Wm}^{-2}$ desde 1750. Este valor es, por mucho, el mayor de todos los posibles forzamientos asociados a los diferentes mecanismos responsables de los cambios de clima que se analizan en este apartado.

Esencialmente hasta aquí se han presentado los motores del clima que están relacionados con el balance de radiación a escala global. Como ya se indicó, el clima es consecuencia de los flujos de energía que se producen entre diferentes partes del Sistema Climático. En ese momento

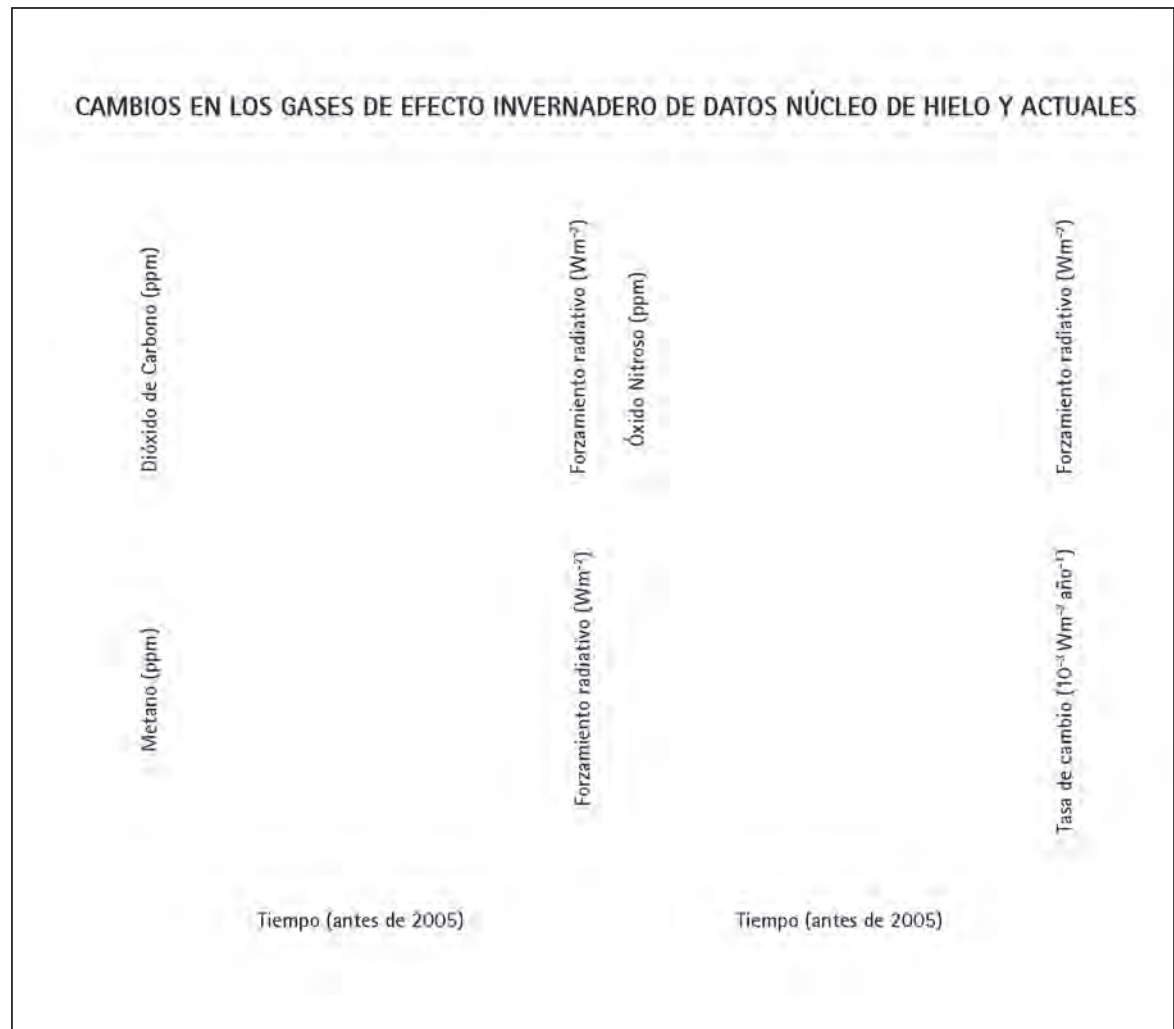


Figura 2. Concentraciones y forzamiento radiativo del (a) dióxido de carbono (CO_2), (b) metano (CH_4), (c) óxido nítrico (N_2O) y (d) la tasa de cambio de su forzamiento radiativo combinado durante los últimos 20.000 años reconstruida a partir del hielo antártico y de Groenlandia y de datos de nieve granular (símbolos) y mediciones atmosféricas directas (paneles a, b, c, líneas rojas). Las barras grises muestran los niveles de variabilidad natural reconstruidos durante los últimos 650.000 años. La tasa de cambio del forzamiento radiativo (panel d, línea negra) se calculó mediante una curva de ajuste spline de los datos de concentración. La resolución temporal varía desde aproximadamente 20 años para lugares con gran acumulación de nieve como Law Dome, Antártida, hasta aproximadamente de 200 años para lugares con poca acumulación como Dome C, Antártida. La flecha muestra el pico en la tasa de cambio del forzamiento radiativo que resultaría si las señales antropogénicas de CO_2 , CH_4 , y N_2O se suavizaran de acuerdo con las condiciones de baja acumulación en el Dome C. La tasa de cambio negativa del forzamiento, aproximadamente en 1600, en el panel de alta resolución insertado en d, es debido a una disminución del CO_2 de unas 10 ppm en el registro del Law Dome (Fuente: IPCC 2007).

es cuando entran en juego una gran cantidad de procesos que tienen su propia dinámica interna, con una gran riqueza de escalas temporales, que hacen que el sistema sea realmente complejo. En consecuencia el Sistema Climático es de muy difícil tratamiento y sólo se puede abordar su estudio completo mediante la simulación numérica. Lo que debe quedar claro es que siempre que se modifique el funcionamiento de una pieza del engranaje, finalmente se producirá un cambio en el clima. Se puede encontrar información mucho más detallada en IPCC (2007).

Cuando hoy día se habla de Cambio Climático (inducido por el hombre) se hace referencia al cambio de clima observado en nuestros días, consecuencia de la intensificación del EI. En última instancia se trata de un cambio

producido en el funcionamiento del planeta Tierra como consecuencia de la actividad humana. Esto es lo que se ha acuñado como Cambio Global y que ha hecho decir a algunos investigadores, entre ellos el nobel Paul Crutzen, que el planeta se encuentra en una nueva era, caracterizada por el impacto antrópico. Por esta razón proponen que sea denominada «Antropoceno» (Crutzen y Stoermer 2000).

Observación de los cambios

Pero, realmente, ¿hay Cambio Climático?, se pregunta mucha gente, y nos preguntan a los especialistas. Desde un punto de vista general la respuesta es sí. Siempre ha habido cambios en el clima del planeta. Y ahora, ¿en el Antropoceno?, pues también. Hay dos aspectos del cambio

de clima actual que conviene remarcar. El primero es que, a diferencia de los anteriores, tiene una escala temporal tan corta que se pueden apreciar los cambios en un tiempo comparable con la vida de una persona. El segundo es que nunca antes la humanidad había tenido capacidad para perturbar el clima a escala global. Se da la circunstancia de que el clima del planeta permitió la vida, incluida la humana, y finalmente la especie humana ha sido capaz de modificar el clima. Estas dos características permiten asegurar que, estrictamente hablando, no existe en los cambios de clima del pasado precedentes del cambio de clima actual.

En este apartado se expondrán algunas de las evidencias del cambio climático actual. En el siguiente se tratará el procedimiento para poder concluir que es la actividad humana la responsable de los cambios observados.

Indica el IPCC en su cuarto, y último, informe (IPCC 2007) que en relación con el tercer informe (se emplean los acrónimos TAR para el tercero y AR4 para el cuarto) se tienen ahora mejores bases de datos, más evidencias, mayor cobertura geográfica y mejor comprensión de las incertidumbres. En consecuencia se indica en AR4 que el calentamiento del Sistema Climático es inequívoco, como se deduce de las observaciones de aumento de las temperaturas medias a escala planetaria del aire y del océano, fusión general de nieve y hielo, y elevación global del nivel medio del mar.

Con respecto a la temperatura media a escala planetaria del aire junto al suelo, en TAR (IPCC, 2001) se había calculado el calentamiento para el intervalo 1901–2000 y se obtuvo una tendencia lineal de $0,6 \pm 0,2^\circ\text{C}$ por siglo, valor que se ve superado por el calculado en AR4 para 1906–2005, de $0,74 \pm 0,18^\circ\text{C}$ por siglo. Esta aceleración del calentamiento se pone más claramente de manifiesto si se emplean sólo los últimos cincuenta años de los cien indicados –1956–2005– y más aún con los últimos veinticinco. En estos casos la tendencia lineal resultante es $1,28 \pm 0,26^\circ\text{C}$ por siglo y $1,77 \pm 0,52^\circ\text{C}$ por siglo, respectivamente.³ Muy probablemente los aumentos de temperatura reseñados no tienen precedente en la Tierra durante los últimos 16.000 años, por lo menos.

Se han observado asimismo cambios en los extremos de temperatura, consistentes con el calentamiento en las capas bajas de la atmósfera. Así, ha disminuido el número de noches frías y con escarcha, mientras que ha aumentado la frecuencia de días y noches cálidos y de olas de calor.

Si se analizan la distribución espacial de las tendencias –mayores sobre tierra que sobre los océanos– y los valores estacionales se encuentran diferencias importantes. Lo mismo ocurre si se calculan las tendencias por separado para temperaturas máximas y mínimas. A modo de ejemplo, se dan los resultados de un análisis de la tendencia de la temperatura en las Islas Baleares a partir de una serie de treinta años, hasta 2006 (OCLIB 2007). Para la temperatura máxima se ha obtenido una tendencia lineal de $4,83 \pm 1,85^\circ\text{C}$ por siglo, siendo para la mínima $5,14 \pm 1,89^\circ\text{C}$ por siglo. Los máximos valores se han encontrado en verano

para la mínima $-8,01 \pm 3,17^\circ\text{C}$ por siglo– y en primavera para la máxima $-7,99 \pm 3,01^\circ\text{C}$ por siglo–. Es importante notar las grandes diferencias encontradas con respecto a los valores globales, incluso con el mayor de los citados anteriormente que corresponde a una serie de veinticinco años.

La temperatura media en el océano también ha aumentado, al menos hasta profundidades de unos 3000 m. Se estima que desde 1955 el océano ha absorbido del orden del 80% del exceso de calor consecuencia de la intensificación del El. Lo anterior da como consecuencia una dilatación del agua marina y una importante contribución a la elevación del nivel del mar.⁴

Además, hay que señalar cambios importantes en la criosfera. Por ejemplo, el hielo marino ártico ha disminuido su superficie por término medio un 2,7% por década, intensificándose la reducción en los veranos del hemisferio norte, en que ha sido del 7,4%. Especialmente notable fue la reducción de superficie cubierta por hielo –por lo menos con un 15% de hielo– en el verano de 2007, una vez elaborado AR4. La superficie cubierta había alcanzado un mínimo en verano de 7,5 millones de km^2 (media en 1979–2000) mientras que en el verano de 2007 se llegó a sólo 4 millones de km^2 , la menor superficie desde que existen satélites de observación de la Tierra. Los valores para el verano de 2008 muestran una ligera recuperación con respecto a las de 2007, pero muy por debajo de la media anteriormente indicada.⁵

En la fig. 3 se resumen los cambios observados en el último siglo y medio, para el valor medio global de la temperatura en superficie (panel a), para el nivel medio del mar (panel b) y para la superficie del hemisferio norte cubierta de nieve (panel c). En la escala relativa de la izquierda de la fig. 3 se da la variación de dichos cambios con respecto al valor medio 1961–1990.

El régimen pluviométrico mundial también se está viendo afectado por el cambio climático actual. En primer lugar hay que decir que se ha producido un incremento continuo del contenido total de vapor de agua en la atmósfera, coherente con el aumento de temperatura en la troposfera. La precipitación se ha visto desigualmente modificada en diferentes áreas geográficas. Mientras ha aumentado de forma significativa en zonas orientales de América del Norte y del Sur, Europa septentrional y Asia septentrional y central, el clima es ahora más seco en el Sahel, Mediterráneo, África meridional y parte de Asia meridional. Si se atiende a los extremos, por una parte, la ocurrencia de episodios de lluvias fuertes ha aumentado de frecuencia sobre la mayor parte de áreas sobre tierra, pero, por otra, se han observado sequías más intensas y duraderas desde la década de los 70, particularmente en trópicos y subtrópicos, a veces combinadas con inundaciones en las mismas zonas geográficas.

Es difícil obtener tendencias globales para la precipitación debido, sobre todo, a la característica discontinuidad de la variable y a los métodos de medida. Como ejemplo de escala mucho menor, se dan a continuación resultados de

3 El calentamiento se ha observado en la temperatura media global en superficie y en la troposfera. En niveles superiores, por ejemplo, en la estratosfera se ha observado un enfriamiento desde 1979 entre $0,3^\circ\text{C}$ y $0,9^\circ\text{C}$ por década, aunque se ha reducido en los últimos años.

4 La variación del nivel del mar es un problema complejo, fuera del alcance de este trabajo. Desde un punto de vista climático las principales contribuciones, casi por igual, se deben a la dilatación del agua marina –incluyendo el efecto de la salinidad– y a la fusión de los hielos continentales. En escalas de tiempo geológicas se han producido cambios del nivel del mar muy importantes; por ejemplo, se estima que en las glaciaciones el mar se encontraba más de 100 m por debajo del nivel actual.

5 Información extraída de <http://nsidc.org/arcticseaicenews/index.html>.

un análisis de la tendencia de la precipitación en las Islas Baleares a partir de una serie de cincuenta y cinco años, hasta 2006 (OCLIB 2007). Suavizando la serie de precipitación anual con un filtro de cinco años, se ha obtenido una tendencia de -170 ± 123 mm por siglo, que se convierte en -192 ± 38 mm por siglo si se filtra la serie anual con un promediado de treinta años. Hay que tener en cuenta que la precipitación normal en Islas Baleares está próxima a los 600 mm por año, lo que representa una disminución de la precipitación con una tendencia del orden de un 30% en cien años. Esta reducción no se ha dado por igual en todas las estaciones del año ni para todos los tipos de precipitación. Las disminuciones han sido mayores en otoño e invierno y mucho menores en primavera y verano, ligadas a una disminución del número de días con lluvias moderadas, aunque ha aumentado el número de días con lluvias débiles y, en menor medida, los de lluvias fuertes.

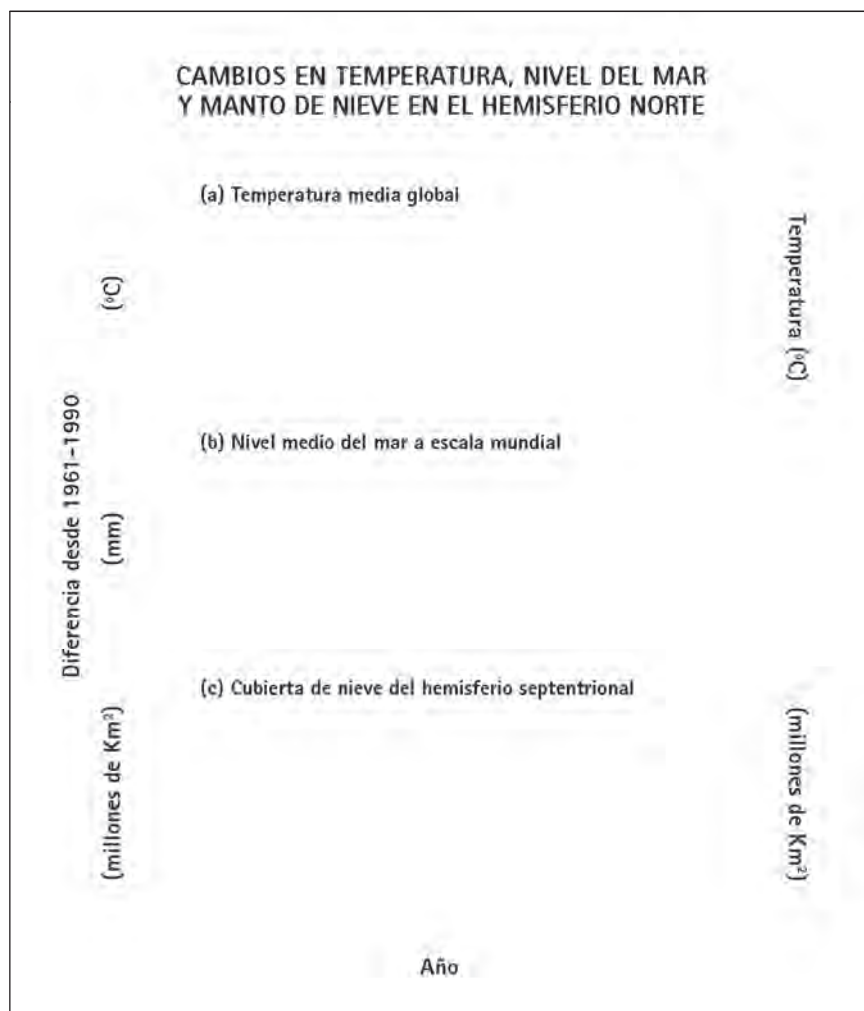


Figura 3. Cambios observados en (a) la temperatura media de la superficie mundial, (b) el promedio del nivel del mar mundial según datos de mareógrafos (azul) y de satélites (rojo) y (c) la cubierta de nieve del hemisferio septentrional en marzo-abril. Todos los cambios son respecto a los promedios correspondientes al periodo 1961–1990. Las curvas suavizadas representan los valores promedio por decenio mientras que los círculos muestran los valores anuales. Las áreas sombreadas indican los intervalos de incertidumbre estimados por un análisis integral de las incertidumbres conocidas (a y b) y de las series de tiempo (c) (Fuente: IPCC 2007).

Los cambios observados en el régimen pluviométrico se explican, en parte, por el aumento indicado del contenido de vapor de agua en el aire, pero también por el cambio que se ha producido en patrones de circulación atmosférica característicos de la variabilidad natural del clima, entre otros la Oscilación del Atlántico Norte –NAO, de su nombre en inglés– y el fenómeno de El Niño/Oscilación Austral –ENSO, de El Niño/Southern Oscillation.

De algunos otros fenómenos extremos, no comentados, se tiene confianza en los cambios observados, por ejemplo, aumento en el número e intensidad de los ciclones tropicales atlánticos, pero de otros –tornados, rayos, granizo, hielo marino antártico y tormentas de polvo– no se tiene por el momento confianza suficiente en los resultados para asegurar que hayan experimentado variación en el clima presente.

Para mayor información sobre los cambios observados es imprescindible consultar AR4 (IPCC 2007).

Atribución del cambio de clima observado

Con el término atribución se quiere indicar el proceso por el cual se evalúa si los cambios observados son consistentes con respuestas cuantitativas a las diferentes causas de cambio en el clima del planeta, simuladas mediante modelos bien probados, y no lo son con explicaciones alternativas físicamente posibles. En este apartado se dará por cierto que se puede simular de forma suficientemente adecuada el clima; en el próximo se tratará de dar argumentos para dejar claro que eso se puede hacer.

Desde que el IPCC elaboró su primer informe, en 1990, se ha tratado el tema de la atribución. En el primer informe (FAR) no había suficientes evidencias observacionales del efecto antrópico sobre el clima. En el segundo (SAR) se concluyó que el conjunto de evidencias sugerían una influencia humana discernible sobre el clima del siglo xx. En TAR se indicó que era probable que la mayor parte del calentamiento observado en los últimos cincuenta años fuera debido al incremento de las concentraciones de los GEI. Desde ese informe, la confianza en la evaluación del efecto de la humanidad en el cambio climático reciente ha aumentado de forma considerable. Se dispone de más evidencias y se ha mejorado la metodología para la atribución. Todo ello aparece en AR4 y se resume a continuación.

Para la atribución del cambio climático actual se emplearán aquí los resultados para la temperatura, que es la variable mejor determinada y cuya simulación está mejor resuelta. Se trata de comparar la evolución observada de la temperatura con aquella que los modelos son capaces de simular. En la fig. 4 se pueden ver los resultados de la comparación de la temperatura media global con la que simulan, en diferentes circunstancias, los modelos numéricos de clima para el siglo xx. Tanto en el panel a como en el b, la curva negra representa la evolución de la temperatura media en superficie para todo el planeta. Los valores numéricos deducidos de la escala de la izquierda

da son las diferencias de temperatura con respecto a la media en el periodo 1901-1950. La curva roja del panel a representa la evolución media de la temperatura simulada. Para obtenerla se promedian los resultados de cada uno de los modelos individuales, cuyas diferentes realizaciones aparecen representadas en ocre. Para esta simulación los modelos incluyen las causas conocidas del cambio de clima; en concreto las naturales, con inclusión de las erupciones volcánicas y aquellas que son consecuencia de la actividad humana, por medio de las evoluciones conocidas de las concentraciones de GEI en la atmósfera y también de aerosoles. El resultado de este experimento de atribución se puede resumir diciendo que existe una buena correlación entre las evoluciones de la temperatura observada y simulada, que la envolvente de las simulaciones individuales incluye casi por completo la curva de observaciones y que la media de los modelos sería una buena aproximación de la observación, convenientemente filtrada por un promedio temporal (no mostrado en la figura).

En el panel b se presenta el resultado de simular la evolución de la temperatura pero sólo empleando las causas naturales del cambio de clima. Como antes, se representan las realizaciones individuales de los modelos, en azul claro, y la media de todas las simulaciones, en azul más oscuro. En este caso no se pueden sacar las mismas conclusiones. Los forzamientos naturales sólo pueden explicar la evolución de temperatura aproximadamente hasta mitad del siglo pasado. De hecho, si se comparan los dos paneles no se observan grandes diferencias entre ambas simulaciones durante ese intervalo. Las diferencias aparecen en la segunda mitad del siglo xx. Es necesario introducir en las simulaciones las causas antrópicas para poder explicar la tendencia de la temperatura en la segunda mitad.

En TAR (IPCC, 2001) ya se habían realizado este tipo de experimentos pero las conclusiones no eran de tanta confianza como en el AR4. Además, ahora se han realizado estudios equivalentes para los diferentes continentes, para tierras y océanos por separado y para otras variables distintas de la temperatura. Los resultados han sido coherentes con lo anteriormente expuesto.

La investigación climática debe tender siempre a reducir incertidumbres y paralelamente a conseguir que las simulaciones sean más realistas. Si se observa la figura 4, existe una discrepancia importante entre las simulaciones y la temperatura media en superficie, calculada a partir de medidas directas, en torno a 1940. Thompson et al. (2008) han analizado el origen de las observaciones de temperatura y concluyen que existe un sesgo en los valores observados como consecuencia del método de medida de la temperatura superficial del mar que, obviamente, forman parte de la temperatura superficial del planeta. Si se corrigieran los valores observados se reduciría la discrepancia, acercándose la evolución observada de la temperatura a la simulada. En el momento de hacerse público

AR4 lo anterior no se conocía, pero aún así se consideraban los resultados suficientemente realistas para indicar que «la mayor parte del aumento observado de la temperatura media global desde la mitad del siglo xx es muy probable⁶ que sea consecuencia del incremento observado de la concentración de GEI antropogénicos».

Simulación del clima de la Tierra mediante modelos numéricos

El conocimiento de los mecanismos que determinan el clima, que fueron expuestos en el apartado 2, es parcial pero suficiente para poderlo simular (¡por supuesto, no en el laboratorio!, pero sí mediante complejos modelos ejecutados en potentes ordenadores). Se ha podido reproducir con garantías suficientes el clima actual y el del pasado más reciente, así como rasgos fundamentales conocidos del clima en eras geológicas pretéritas. Gracias a esto se han podido realizar ejercicios de atribución, como se ha indicado en el apartado 4, y se puede pensar también en inferir las posibilidades de clima futuro, incluyendo el papel del hombre. Esto último será tratado en el siguiente apartado.

Veamos ahora con un cierto detalle qué son los modelos que permiten la simulación del clima. Hay que indicar, en primer lugar, que los modelos no son una invención de los investigadores del clima; en física, y en otras ciencias, se emplean modelos con mucha generalidad y han resultado ser extraordinariamente útiles para el avance del conocimiento. En términos generales, un modelo es una simplificación de la realidad que se utiliza como herramienta para describir y explicar fenómenos de interés científico. A veces se construyen los modelos mediante ecuaciones matemáticas que resumen relaciones empíricas entre variables características del sistema objeto de estudio. Por ejemplo, se pueden obtener esas relaciones a partir de un adecuado tratamiento estadístico de las variables. Otras veces son las leyes de la física, previa e independientemente establecidas, las que proporcionan la relación entre las variables. En este caso, además, permiten interpretar el porqué de esa relación ya que, de hecho, es lo que estas leyes expresan. Finalmente también son ecuaciones matemáticas las que relacionan las variables, pero ahora basadas en leyes físicas.

Sea como fuere, se dispone de un conjunto de ecuaciones matemáticas que permite describir de forma aproximada –no se olvide que se trata de una simplificación– la realidad. Es precisamente este hecho el que va a permitir explicar, al menos en parte, las discrepancias que aparezcan entre una descripción simulada de la realidad mediante un modelo y la realidad de la observación de un fenómeno real.

Una vez se dispone del conjunto de ecuaciones que constituyen un modelo hay que escribirlas de forma adecuada para poder obtener información cuantitativa referida al sistema estudiado. En el caso que nos ocupa, habría que obtener, por lo menos, los valores de temperatura y precipitación para conocer los rasgos fundamentales del clima. Pero, además, habría que hacerlo en todo el pla-

⁶ Indica con este término el IPCC que la probabilidad supera el 90%.

neta y, en realidad, a diferentes niveles de la atmósfera, desde los más bajos, en contacto con el suelo o los mares, a los más altos. Y esto sólo por lo que hace referencia a la atmósfera porque, en los otros subsistemas, haría falta conocer otras muchas variables (a modo de ejemplo, salinidad y temperatura en los océanos, masa de hielo, propiedades de los suelos y de la vegetación) y también a diferentes niveles o profundidades. La conclusión que se debe sacar de lo anterior es que es necesario aplicar las ecuaciones del modelo a una gran cantidad de puntos del espacio. Son muchas las operaciones matemáticas a realizar para determinar todas las variables que describen

el estado del Sistema Climático en un solo instante; pero para caracterizar al clima debe conocerse lo que ocurre, no en un instante concreto, sino a lo largo de intervalos de tiempo suficientemente largos, es decir, formados por una enorme sucesión de instantes individuales.

¿Cómo se puede abordar esta tremenda tarea? La contestación no es inmediata. En primer lugar, si se quiere obtener información climática útil en un tiempo razonable, se deben utilizar ordenadores muy potentes, los más potentes del mundo. Para ello hay que volver a simplificar el modelo, escribiéndolo en una forma que resulte adecuada para trabajar con ordenador. Una vez hecho esto, serían las máquinas las encargadas de realizar los millones y millones de operaciones numéricas necesarias para obtener en un tiempo razonable simulaciones del clima para varias décadas, siglos, etc. Muchas veces se habla de simulaciones numéricas del clima para recoger el modo en que se obtiene la información climática que se desea.

Los modelos más avanzados de simulación climática incluyen: el tratamiento de procesos atmosféricos, oceánicos, en la superficie terrestre y en la criosfera, la química atmosférica y la modelización de aerosoles. También tratan de forma acoplada las interacciones atmósfera-oceano. Algunos de los modelos disponen de mecanismos para controlar los flujos de energía en valores razonables, pero hoy día, debido a los avances en la investigación, la mayoría de ellos no necesitan este ajuste pues los flujos obtenidos directamente en la simulación son ya realistas. Se utiliza el nombre genérico de Modelos de Circulación General Atmósfera-Océano –AOGCM, acrónimo correspondiente a su denominación en inglés– para referirse a los modelos de simulación del clima que incorporan ecuaciones para el tratamiento de los procesos mencionados. Los muchos modelos que existen, normalmente vinculados a los centros de investigación punteros del mundo, dan resultados de simulación climática diferentes, aunque todos ellos plausibles. Existen campañas de intercomparación de resultados para verificar su comportamiento que, además, permiten establecer niveles de confianza para los resultados. El IPCC basa precisamente en las simulaciones gran parte de sus informes de evaluación (véase, p.e., capítulos 8 y 9 de AR4, IPCC 2007). La confianza en la simulación numérica del clima se ha alcanzado verificando que los resultados son suficientemente realistas cuando se les compara con las observaciones. Estos resultados afectan a los diferentes subsistemas del Sistema Climático y a modos conocidos de variabilidad del clima actual, entre ellos el fenómeno de El Niño/Oscilación Austral (ENSO), la Oscilación del Atlántico Norte (NAO), las situaciones de bloqueo anticiclónico y la variabilidad de los monzones. Sin embargo, no es la verificación frente al clima presente la única fuente de confianza. Desde un punto de vista conceptual la fuente primera es el uso que se hace en los modelos de leyes físicas establecidas previa e independientemente del problema de la simulación del

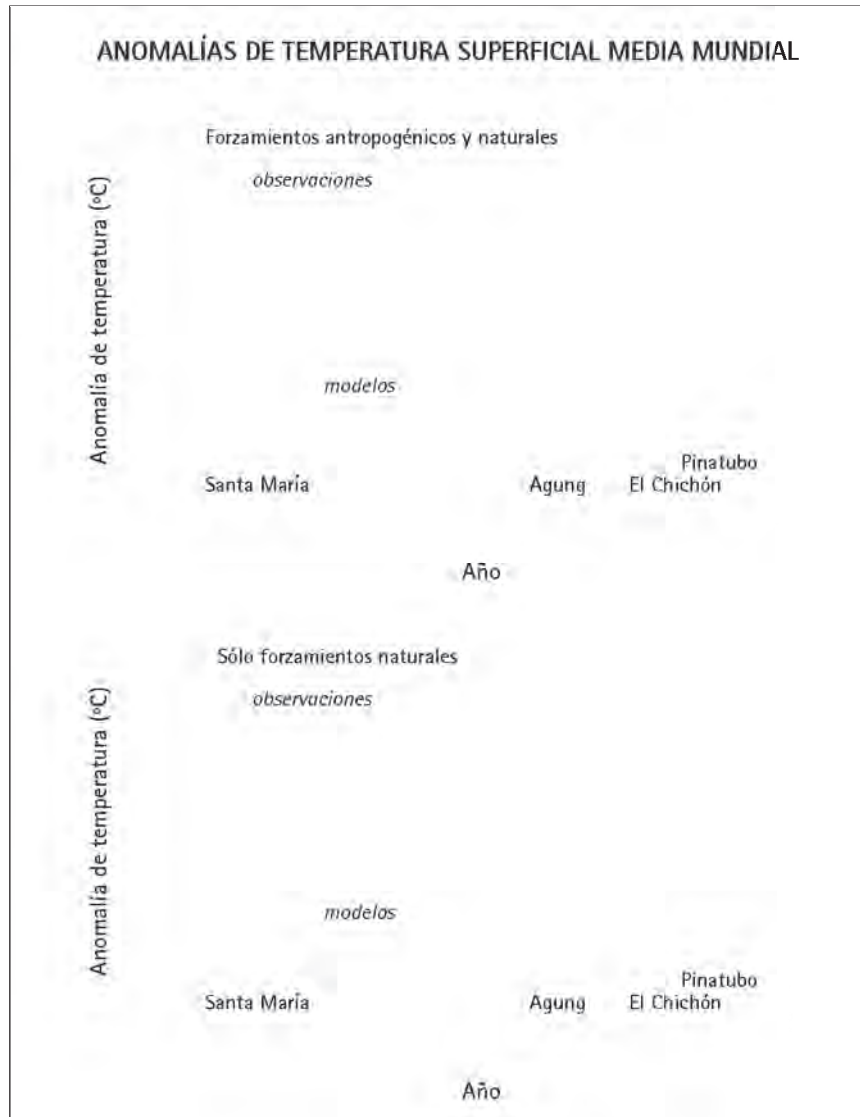


Figura 4. (a) Anomalías en la temperatura superficial media mundial relativas al periodo de 1901 a 1950, según observaciones, (línea negra) y como resultado de simulaciones con forzamientos antropogénicos y naturales. La curva gruesa, en rojo, muestra la media del conjunto de varios modelos y cada línea delgada ocre muestra una simulación individual. Las líneas grises verticales indican importantes fenómenos volcánicos. (b) Igual que en (a), sólo que las anomalías simuladas de la temperatura media mundial son solamente para los forzamientos naturales. La curva gruesa azul muestra la media del conjunto de modelos y cada curva azul más clara muestra una simulación individual. Cada simulación fue realizada para que la cobertura temporal correspondiera con la de las observaciones (Fuente: IPCC 2007).

clima. Además, ha sido posible simular rasgos importantes del clima de los últimos 2.000 años y de cambios climáticos anteriores, como el periodo cálido en el Holoceno hace 6.000 años y la variabilidad en las glaciaciones. Ni que decir tiene que los resultados son adecuados para tener confianza en el uso de los modelos pese a que todavía permanecen fuentes de incertidumbre.

Una de las principales ventajas de la simulación del clima mediante modelos es que se pueden activar o desactivar a voluntad procesos incluidos en el modelo. Para ello se eliminan del conjunto de ecuaciones del modelo aquellas que afectan al proceso en cuestión. El modelo es capaz entonces de simular el clima del planeta actuando o no el proceso (o procesos) bajo estudio. De esta forma, por ejemplo, se puede incluir, tras una erupción volcánica, el efecto adicional de los aerosoles expulsados o se puede eliminar la intensificación del EI y considerar concentraciones preindustriales de GEI. En esto precisamente está basada la atribución del cambio climático tratada en el anterior apartado.

Si no se quiere, o no se puede, recurrir a grandes ordenadores, también hay soluciones más modestas, pero no por eso menos útiles. Se puede acceder a una segunda vía de simulación del clima por medio de una nueva simplificación del Sistema Climático. Es decir, simplificar la complejidad del modelo, que ya era de por sí una simplificación de la realidad, de forma que se pueda trabajar con él en ordenadores de tipo personal o similares. Lo que se trata

entonces de conseguir con los modelos sencillos es que sus simulaciones sean compatibles con aquellas que se realizan con los AOGCM.

Para hacerse una idea, en el máximo extremo de la sencillez, se podría considerar la Tierra como una esfera que recibiera energía del Sol y que mantuviera el equilibrio de esa energía con la energía que se refleja y la que la propia Tierra disipa hacia el espacio. En estas condiciones se determina una temperatura, llamada de equilibrio, que resulta ser aproximadamente -18°C y que es muy diferente de la temperatura media en la Tierra, unos 15°C . Anteriormente se han apuntado esas mismas cifras al introducir el EI natural. O sea, la temperatura de equilibrio se obtiene simplificando al máximo el sistema –en concreto prescindiendo de la atmósfera–, con lo que esas condiciones se parecen más a las de la Luna que a las de la Tierra. La toma en consideración de la atmósfera permite asignar al EI un aumento de temperatura de unos 33°C lo cual, si se piensa bien, es espectacular. Sirva como comparación, por ejemplo, que se piensa que la oscilación de temperatura asociada a las eras geológicas o a cambios climáticos abruptos no llegan ni a la mitad de lo indicado para el calentamiento debido al EI natural (Masson-Delmotte et al. 2005).

Con otros modelos sencillos, por supuesto no tanto como el anterior, es posible calcular la distribución de la temperatura de equilibrio para diferentes latitudes de la Tierra, considerar de forma elemental el papel de las nubes, determinar otros climas potenciales con todo el hielo fundido o con la Tierra totalmente cubierta de hielo, las transiciones entre ambos, etc. Una ventaja de los modelos sencillos sobre los más complejos es que se pueden realizar un gran número de experimentos diferentes, cambiando algunas de las condiciones de la simulación, ya que necesitan mucho menos tiempo para resolver las ecuaciones que los modelos más complejos.

Proyección del clima hacia el futuro

Es importante insistir que los modelos climáticos son la herramienta más importante, si no única, para llevar a cabo simulaciones del clima del planeta. Para poder utilizarlos con garantía se han llevado a cabo experimentos para reproducir el clima presente, el clima del pasado y dar explicación al cambio de clima que la Tierra está experimentando. Como las ecuaciones de partida provienen de leyes físicas y la simulación es realista se tiene una gran confianza en el uso de modelos. Es evidente que quedan aspectos por conocer del funcionamiento del Sistema Climático y que ese desconocimiento genera incertidumbre. Sin embargo, aceptando los resultados de la simulación, al ser verificados mediante la observación, estamos indicando que el conocimiento que se tiene sobre el funcionamiento del sistema es suficiente y que lo que se desconoce no sería capaz de modificar de forma sustancial las simulaciones. Si no fuera así, o sea, si nuestra

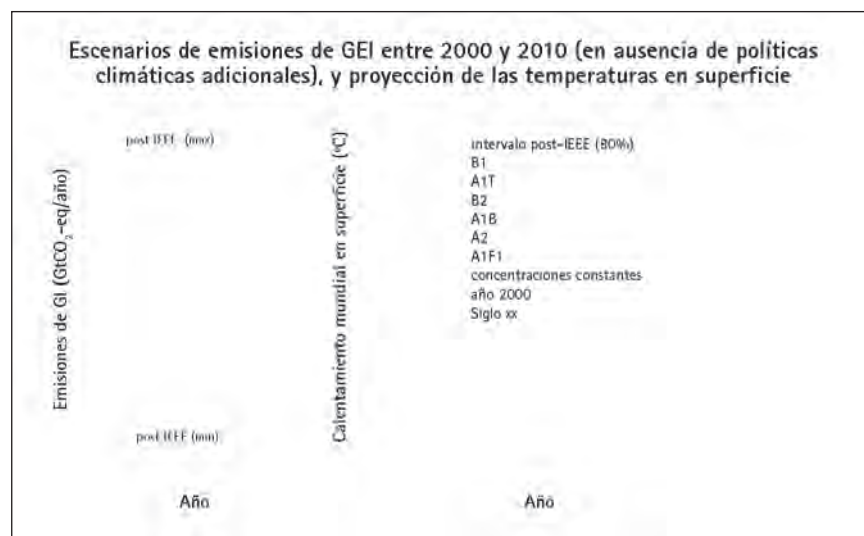


Figura 5. Panel izquierdo: emisiones mundiales de GEI ($\text{CO}_2\text{-eq}$) en ausencia de políticas climáticas: seis escenarios SRES ilustrativos (líneas de color), junto con el percentil del 80% de escenarios recientes publicados desde el SRES (post SRES) (área sombreada en gris). Las bandas de color a la derecha representan la totalidad de los escenarios post IEEE. Las emisiones abarcan los gases CO_2 , CH_4 , N_2O y F. Panel derecho: las líneas continuas representan promedios mundiales multimodelo del calentamiento en superficie para los escenarios A2, A1B y B1, representados como continuación de las simulaciones del siglo xx. Estas proyecciones reflejan también las emisiones de GEI y aerosoles de corta permanencia. La línea rosa no es un escenario, sino que corresponde a simulaciones de AOGCM en que las concentraciones atmosféricas se mantienen constantes en los valores del año 2000. Las barras de la derecha indican la estimación óptima (línea gruesa dentro de cada barra) y el intervalo probable evaluado para los seis escenarios SRES considerados en el periodo 2090-2099. Todas las temperaturas son anomalías respecto del periodo 1980-1999 (Fuente: IPCC 2007).

ignorancia implicara consecuencias de importancia en las simulaciones, la investigación ya lo habría detectado.

Dicho esto, debe quedar claro que la simulación del clima presente no es el mismo problema que la simulación del clima futuro. En el primer caso se conocen los cambios que se han producido en el pasado para llegar desde el pasado al presente. Se sabe cómo ha cambiado la radiación interceptada por la Tierra y se conocen los cambios en la composición atmosférica. No sólo por lo que afecta a la concentración de los GEI sino también, por ejemplo, a las erupciones volcánicas. El forzamiento de los modelos con las condiciones reales conocidas ha permitido reconstruir el clima presente. Pero a partir del presente no se sabe cuáles van a ser las condiciones en la atmósfera terrestre y sin embargo su conocimiento es imprescindible para poder simular el clima futuro.

Sabemos, por ejemplo, que las emisiones anuales de CO₂ de origen fósil han aumentado desde una media de 6,4 GtC⁷ por año en la década de los 90 a 7,2 GtC por año en el periodo 2000-2005. Estas emisiones, junto con las del pasado, han determinado, en parte, la concentración de CO₂ en la atmósfera, lo mismo que otros procesos lo han hecho con la de otros GEI. El problema de determi-

nar la concentración de GEI a partir de las emisiones no es sencillo; hay que recurrir nuevamente a la simulación mediante modelos, en este caso, del ciclo del carbono y de otros elementos. Hay que tener en cuenta, por ejemplo, la fijación de carbono en el suelo y en los mares (sumideros), lo que a su vez depende de muchos factores.

Supuesto que esté resuelto este problema, queda por conocer la evolución hacia el futuro de las emisiones de GEI. Lo que sí debe estar claro es que depende de muchos condicionantes, difíciles de determinar, fundamentalmente de carácter socioeconómico. Lo que se hace es trabajar con diferentes hipótesis plausibles que reciben el nombre de escenarios. Desde los primeros informes del IPCC (FAR y SAR) se ha prestado atención a la definición de escenarios de emisiones, que en un principio estaban incluidos en los propios informes. Sin embargo, tras el segundo se encargó un trabajo específico sobre escenarios (IPCC, 2000) que dio lugar a los que actualmente se usan para proyectar el clima hacia el futuro. Reciben el nombre de SRES, acrónimo que recoge el carácter y título de la obra: Special Report on Emissions Scenarios.

En resumen, se trabaja con cuatro líneas evolutivas (A1, A2, B1 y B2) condicionadas por «fuerzas» como población, economía, tecnología, energía, agricultura y usos del suelo. En A1 y A2 se da más peso al crecimiento económico, mientras que en B1 y B2 predominan los aspectos ambientales. Por otra parte en A1 y B1 se tiende a un mundo globalizado, en tanto que en A2 y B2 se enfatiza en soluciones regionales y locales. Cada una de estas líneas da lugar a diferentes escenarios, hasta completar un total de cuarenta. Normalmente se utilizan familias, coincidentes en nombre con las líneas, con excepción del A1 que se desglosa en tres:

- A1FI, con uso intensivo de combustibles fósiles.
- A1T, con uso de fuentes de energía no fósil.
- A1B, con uso equilibrado de diferentes fuentes.

Es evidente que no se sabe el camino que tomará la humanidad a partir de ahora, en consecuencia, todos los escenarios se consideran igualmente probables.

Cada uno de los escenarios de emisiones SRES lleva asociados valores concretos de emisiones de GEI a lo largo del siglo XXI. Entonces, mediante el uso de modelos adecuados se deducen las concentraciones futuras de GEI y con la evolución futura de dichas concentraciones se puede proyectar el clima hacia el futuro, gracias a los modelos de simulación del clima. El resultado es un conjunto de proyecciones climáticas, para cada uno de los SRES considerados, que, por diferencia con unas condiciones climáticas de referencia, da lugar a diferentes escenarios futuros de cambio climático. Los escenarios o proyecciones pueden ser globales o estar restringidos a regiones concretas de la geografía mundial.

En el panel izquierdo de la fig. 5 se muestra la evolución de las emisiones de GEI durante el siglo XXI. En la figura aparecen englobadas las emisiones de todos los GEI

7
GtC: gigatoneladas de carbono, mil millones de toneladas de carbono.

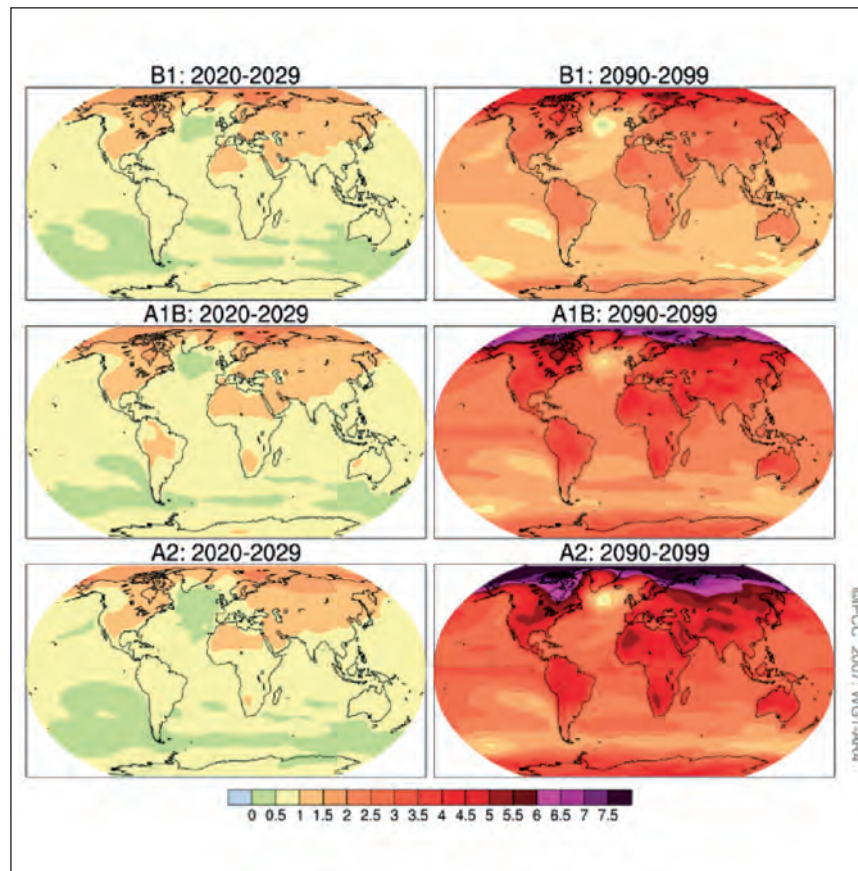


Figura 6. Cambios de temperatura superficial proyectados para inicios y finales del siglo XXI relativos al periodo 1980-1999. Los paneles a la izquierda y a la derecha muestran las proyecciones medias de multimodelos AOGCM para la media por decenios de los escenarios B1 (arriba), A1B (centro) y A2 (abajo) de 2020 a 2029 (izquierda) y de 2090 a 2099 (derecha) (Fuente: IPCC, 2007).

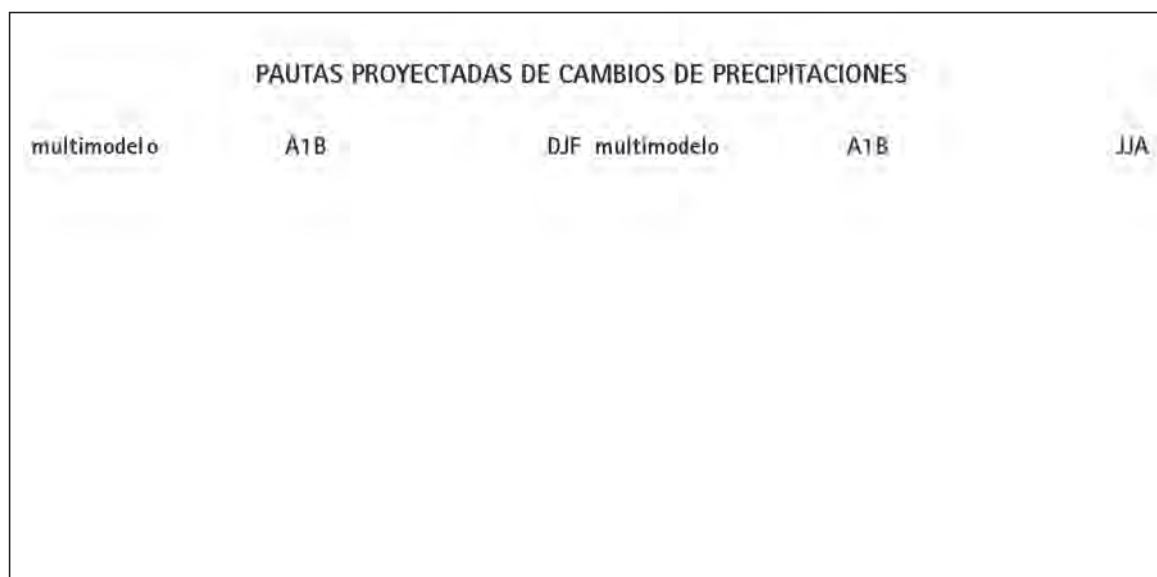


Figura 7. Cambios de la precipitación relativos (en valores porcentuales) para el periodo 2090-2099, respecto del periodo 1980-1999. Los valores son promedios multimodelo basados en el escenario A1B para los periodos diciembre-febrero (izquierda) y junio-agosto (derecha). Las áreas en blanco representan los lugares en que más de un 66% de los modelos coinciden en el signo del cambio, y las áreas punteadas representan los lugares en que más de un 90% de los modelos concuerdan en el signo del cambio (Fuente: IPCC, 2007).

en lo que se denomina CO₂ equivalente, calculado teniendo en cuenta el mismo efecto de intensificación del El que todos los GEI considerados. Además de los escenarios SRES descritos anteriormente se dan aquí resultados para otros escenarios posteriores a la publicación del Informe Especial (IPCC 2000), que modificaban la contribución de ciertas «fuerzas» condicionantes de las líneas evolutivas consideradas. En el panel derecho se muestran proyecciones de la temperatura media en superficie para varias familias de escenarios y la proyección correspondiente a no aumentar la concentración de GEI en los valores del año 2000. Hay que hacer notar que a pesar de este hecho, la temperatura seguiría aumentando aunque, evidentemente, a un ritmo mucho menor.

Si se analizan las proyecciones para las dos primeras décadas de este siglo, los resultados son muy poco dependientes del escenario considerado y del modelo usado (0,2°C por década). Sin embargo, para las décadas finales del siglo no es así, dependen fuertemente del escenario considerado y también del modelo empleado. Por ejemplo, la estimación media multimodelo para el escenario B1 a final de siglo es 1,8°C (probablemente con rango de 1,1°C a 2,9°C) y para el escenario A1FI es 4,0°C (probablemente con rango de 2,4°C a 6,4°C), siempre por encima de la media del periodo 1980-1999. Nótese que esos valores están muy por encima de los observados para el aumento de la temperatura media en superficie para el siglo xx.

Teniendo en cuentas estas proyecciones de temperatura se ha evaluado el efecto sobre el nivel global medio del mar—incluyen también las contribuciones de la fusión del hielo en Groenlandia y Antártida—. La elevación a final del siglo XXI que depende, lógicamente, del escenario considera-

do iría del mínimo 0,18 m a 0,38 m para el escenario B1 al máximo 0,26 m a 0,59 m para el escenario A1FI. Estos valores son relativos al nivel medio del mar en 1980-1999.

Los modelos AOGCM permiten llevar a cabo proyecciones climáticas globales en las que se puede apreciar la variabilidad espacial y temporal. En AR4 se incluyen una gran cantidad de proyecciones (véase IPCC, 2007, capítulo 10) de las que aquí se presentan sólo algunas. En la fig. 6 se tienen mapas de proyecciones medias multimodelo de temperatura superficial para diferentes alcances temporales y escenarios. Conviene destacar dos aspectos; primero, la poca diferencia existente en los mapas de la columna izquierda, como ya se había adelantado; segundo, la desigual distribución geográfica de los aumentos de temperatura, con un claro predominio de los valores en la región ártica donde la temperatura podría aumentar a final de siglo más de 7°C. En general, el calentamiento proyectado para el siglo XXI se espera que sea mayor sobre tierra y a altas latitudes del hemisferio norte y menor sobre el océano Austral y parte del Atlántico norte.

En la figura 7 se tienen proyecciones estacionales de la precipitación. Aunque se estima que en valor global su valor aumente, probablemente en la mayor parte de las regiones subtropicales terrestres decrezca mientras que en latitudes altas muy posiblemente la precipitación será más elevada.

Se han obtenido también proyecciones para otros aspectos importantes del clima. En general se puede decir que todos ellos continúan con la tendencia observada en el siglo xx pero, en la mayor parte de los casos, acentuándose.

Mención especial merece la fusión de los hielos en Groenlandia aunque la escala temporal sea superior al siglo. Hace unos 125.000 años la temperatura en la zona del Atlántico norte se mantuvo durante un perio-

do duradero por encima de las temperaturas actuales. La reducción de la masa de hielo hizo que el nivel del mar se elevara de 4 a 6 m. Pues bien, si la temperatura fuera entre 1,9 y 4,6°C superior a la preindustrial durante al menos mil años, la fusión del hielo de Groenlandia podría producir una elevación del nivel del mar en el planeta de 7 m.

Una de las aplicaciones más importantes de las proyecciones climáticas es el análisis de las consecuencias del cambio climático o, como se conoce habitualmente, de los impactos del cambio climático, lo que tiene una gran incidencia social pues sus efectos son locales. Para poderlos determinar es preciso disponer de proyecciones climáticas con mucha mayor resolución que la que proporcionan los modelos globales. Existen para ello diferentes metodologías que reciben el nombre general de *downscaling*. En una de las más usadas se utilizan modelos de simulación de escala regional anidados en modelos globales que se ejecutan de forma acoplada y simultánea. Se trata del *downscaling* dinámico. Otra posibilidad es emplear relaciones estadísticas empíricas determinadas para el clima presente, que se supone siguen siendo válidas para el futuro, para ganar resolución a partir de proyecciones de clima futuro obtenidas con AOGCM. También existen metodologías que combinan las dos anteriores. Se puede obtener más información sobre *downscaling* en el capítulo 11 de AR4 (IPCC 2007).

Conclusiones

Durante el Antropoceno el planeta Tierra está experimentando un cambio de clima que, estrictamente hablando, no tiene precedente en el pasado. La quema de combustibles fósiles y la actividad humana en general ha modificado la composición de la atmósfera aumentando la concentración de los GEI hasta unos valores nunca alcanzados, al menos, en los 800.000 últimos años. El EI, que ha permitido la vida sobre la Tierra, está siendo intensificado antropogénicamente y está originando un incremento en la temperatura media mundial en superficie durante el siglo xx que no tiene antecedentes, por lo menos, en los últimos 16.000 años. Junto con este cambio de temperatura también se ha observado una elevación del nivel del mar y la reducción de la cobertura de nieve en los conti-

entes y del hielo marino en el océano Ártico. Además se están modificando patrones climáticos, como el régimen de precipitación, la NAO y el fenómeno ENSO, entre otros, y están cambiando las frecuencias de ocurrencia de algunos fenómenos extremos.

De continuar las emisiones de GEI al ritmo actual, el cambio climático observado se verá acelerado en el siglo presente. Incluso, si se mantuvieran las concentraciones de dichos gases en los valores actuales, el aumento de temperatura y los efectos consiguientes se seguirían produciendo, aunque con menor intensidad, durante décadas.

Las consecuencias sociales y económicas de los cambios observados ya empiezan a ser importantes en algunas zonas –cambios de hábitats, agotamiento de la capacidad de adaptación de algunas especies, modificación de periodos de cultivo, problemas con los recursos hídricos, cambio en la distribución y ocurrencia de algunas enfermedades, etc.– pero todavía se cree que serán más importantes a medida que se vaya intensificando el calentamiento. Desde el punto de vista humano, las sociedades menos favorecidas, con menor nivel de desarrollo, serán las más vulnerables.

El calentamiento del clima ya no se puede detener, estamos sufriendo ahora las consecuencias de lo que iniciamos con la Revolución Industrial. Es evidente que tenemos que reducir las emisiones, lo cual es intrínsecamente bueno para el medio ambiente en general, pero también debemos hacer un esfuerzo para adaptarnos al clima que viene y para comprender que, además de convivir con un cierto nivel de riesgo, se deberá hacer frente al coste de la adaptación. En cualquier caso, éste será menor que el coste de no hacer nada. Los políticos tienen que jugar su papel y la sociedad el suyo. Y obviamente, como parte de la sociedad que somos, los científicos también. Se debe intensificar la investigación, reduciendo incertidumbres, mejorando las proyecciones climáticas, dando pistas sobre la reducción de vulnerabilidades y los riesgos climáticos, buscando sistemas más eficientes de aprovechamiento energético, sistemas menos contaminantes, etc.

Seguramente tendremos que cambiar ligeramente el modo de vida para que los países en vías de desarrollo puedan alcanzar un nivel de bienestar adecuado. La humanidad del futuro no espera otra cosa de nosotros.

Agradecimientos:

Las sugerencias realizadas por el Prof. C. Ramis al manuscrito han mejorado considerablemente la calidad de este trabajo, por lo que le doy las gracias. Las figuras son reproducción exacta de las que aparecen en los documentos del IPCC y están incluidas aquí con su autorización tácita.

Bibliografía

- Álvarez, L. W., W. Álvarez, F. Asaro y H. V. Michel. «Asteroid Extinction Hypothesis». *Science* 211 (1981): 654-656.
- Arrhenius, S. «On the influence of carbonic acid in the air upon the temperature on the ground». *Philos. Mag.* 41 (1896): 237-276.
- Berger, A. «Milankovitch theory and climate». *Reviews of Geophysics* 26 (1988): 624-657.
- Croll, J. *Climate and Time in Their Geological Relations: A Theory of Secular Changes of the Earth's Climate*. 2ª ed. Nueva York: Appleton, 1890.
- Crutzen, P. y E. F. Stoermer. «The "Anthropocene"». *Global Change Newsletter* 41 (2000): 12-13.
- Eddy, J. A. «The Maunder Minimum». *Science* 192 (1976): 1.189-1.202.
- Hoyt, D. V., K. H. Schatten y E. Nesmes-Ribes. «The hundredth year of Rudolf Wolf's death: Do we have the correct reconstruction of solar activity?». *Geophys. Res. Lett.* 21 (1994): 2.067-2.070.
- IPCC. *Special Report on Emissions Scenarios* [Nakicenovic, N. y R. Swart (eds.)]. Cambridge y Nueva York: Cambridge University Press, 2000.
- IPCC. *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* [Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell y C. A. Johnson (eds.)]. Cambridge y Nueva York: Cambridge University Press, 2001.
- IPCC. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor y H. L. Miller (eds.)]. Cambridge y Nueva York: Cambridge University Press, 2007.
- Kiehl, J. y K. Trenberth. «Earth's annual global mean energy budget». *Bull. Am. Meteorol. Soc.* 78 (1997): 197-206.
- Köhler, P., F. Joos, S. Gerber y R. Knutti. «Simulating changes in vegetation distribution, land carbon storage, and atmospheric CO₂ in response to a collapse of the North Atlantic thermohaline circulation». *Clim. Dyn.* 25 (2005): 689-708.
- Loulergue, L., A. Schilt, R. Spahni, V. Masson-Delmotte, T. Blunier, B. Lemieux, J. M. Barnola, D. Raynaud, T. F. Stocker y J. Chappellaz. «Orbital and millennial-scale features of atmospheric CH₄ over the last 800,000 years». *Nature* 453 (15 de mayo de 2008): 383-386.
- Lüthi, D., M. Le Floch, B. Bereiter, T. Blunier, J.-M. Barnola, U. Siegenthaler, D. Raynaud et al. «High-resolution carbon dioxide concentration record 650,000-800,000 years before present». *Nature* 453 (15 de mayo de 2008): 379-382.
- Masson-Delmotte, V., A. Landais, N. Combourieu-Nebout, U. von Grafenstein, J. Jouzel, N. Caillon, J. Chappellaz, D. Dahl-Jensen, S. J. Johnsen y B. Stenni. «Rapid climate variability during warm and cold periods in polar regions and Europe». *Comptes Rendus Geoscience* 337 (2005): 935-946.
- OCLIB. «Informe de seguimiento del convenio Universitat de les Illes Balears-Conselleria de Medi Ambient del Govern de les Illes Balears, sobre el Observatori del Clima de les Illes Balears». Informe técnico inédito. Grup de Meteorologia, UIB. 2007.
- Ruddiman, W.F., Ed. *Tectonic Uplift and Climate Change*. Nueva York: Plenum Press, 1997.
- Thompson, D. W., J. J. Kennedy, J. M. Wallace y P.D. Jones. «A large discontinuity in the mid-twentieth century in observed global-mean surface temperature». *Nature* 453 (29 de mayo de 2008): 646-649.
- Wang, Y. M., J. L. Lean y N. R. Sheeley. «Modeling the sun's magnetic field and irradiance since 1713». *Astrophys. J.* 625 (2005): 522-538.
- Yang, F. y M. Schlesinger. «On the surface and atmospheric temperature changes following the 1991 Pinatubo volcanic eruption: a GCM study». *J. Geophys. Res.-Atmos.* 107 (abril de 2002): doi10.1029/2001JD000373.

la economía de la empresa

VICENTE SALAS FUMÁS

Introducción

La empresa es una institución central en el funcionamiento de cualquier sistema económico donde las personas satisfacen sus necesidades a través de la división del trabajo, la colaboración en la producción y el intercambio de bienes y servicios. Como parte del sistema a la empresa le corresponde la función de producir bienes o servicios para su venta en el mercado, una función necesaria para que cada persona pueda hacer compatible la especialización en el trabajo y la satisfacción de sus múltiples necesidades. La empresa se manifiesta en forma de una personalidad jurídica propia y una razón social. La heterogeneidad de empresas en cuanto a dimensión, variedad de bienes o servicios que se venden al mercado, actividades y recursos que se controlan internamente, organización y funcionamiento interno, despierta el interés intelectual de las ciencias sociales en general y de la economía en particular. Por qué existen las empresas, cuál es su naturaleza, cómo se estructuran y funcionan internamente, qué factores influyen en los cambios de naturaleza que se observan en el tiempo, son preguntas que forman parte del inventario de la investigación económica sobre la empresa.¹

Las empresas surgen de las decisiones de personas, empresarios, que además dirigen la asignación de recursos dentro de los límites de su competencia. En empresas complejas la dirección de recursos, conocida genérica-

mente como *management*, debe compartirse entre múltiples especialistas, dando lugar a un ámbito profesional de gran importancia cuantitativa y cualitativa en las sociedades desarrolladas. Paralelamente al estudio y al *conocimiento positivo* sobre la realidad de la empresa, que ha protagonizado la economía y otras ciencias sociales, se ha desarrollado también un *conocimiento normativo* sobre la toma de decisiones y la forma de dirigir la empresa que se enseña como parte de la formación profesional de empresarios y directivos. De la importancia de estas enseñanzas da fe el auge de centros especializados en la formación de directivos profesionales en todo el mundo. Existen, por tanto, dos grandes bases de conocimiento sobre la empresa, con una dinámica de generación y renovación de contenidos que es, a la vez, propia y mutuamente interdependiente: de un lado, la que se vincula al *porqué* de los fenómenos estudiados —análisis positivo— que centra el interés de las ciencias sociales. Su fin último es aprender acerca de las consecuencias de una realidad empresarial u otra sobre el bienestar social. Del otro, la base de conocimiento que gira sobre el *cómo* actuar ante determinados problemas —análisis normativo— del que se ocupan las disciplinas del *management* profesional. El fin del conocimiento creado es ahora contribuir al bienestar particular de quien toma las decisiones en la empresa, en concreto aumentar el beneficio privado.

¹ DiMaggio (2001) y Roberts (2004) ofrecen una visión integrada del pasado reciente y futuro de la empresa desde disciplinas diferentes, sociología y economía, y por tanto complementarias. Véase también Malone et al. (2003) y Salas Fumás (2007).

Resumir el estado de la cuestión sobre el conocimiento positivo y normativo alrededor de la empresa y su gestión es una tarea inalcanzable en un espacio limitado, no sólo por la pluralidad en las preguntas que se plantean sino también por la diversidad de disciplinas académicas que se interesan por ellas. Nos limitaremos, por tanto, a destacar la parte del conocimiento positivo que es atribuible en mayor medida a la investigación económica sobre la teoría de la empresa. De este modo la exposición posterior se organiza como sigue: el primer apartado delimita los antecedentes sobre el campo de interés general de la economía y el encaje que tiene la empresa en ese ámbito. El segundo se ocupa de la investigación sobre los determinantes de las fronteras o límites entre la empresa y el mercado del que forma parte. El tercer apartado se dedica a resumir los avances en la economía de la organización interna de las empresas, mientras que en el apartado cuarto se abordan cuestiones como la persona jurídica de la empresa y las relaciones sociales que integran la economía de la empresa con otras ciencias sociales. En las conclusiones se valoran las aportaciones de la teoría desde la gestión empresarial y sobre las expectativas de la sociedad en cuanto a la buena *performance* de las empresas.

Los antecedentes y marco general de la economía de la empresa

En la llamada *economía de mercado* la relación entre empresas, o entre empresas y consumidores finales, trabajadores, inversores..., se regula a través de los precios que informan del valor relativo de los recursos disponibles en usos alternativos cuando las necesidades a satisfacer exceden a los medios disponibles. La economía de mercado acostumbra a incorporar la institución de la propiedad privada, de manera que el precio es la recompensa monetaria para quien produce y vende lo que otros demandan. Mercado también es sinónimo de libre empresa, que quiere decir igualdad entre los ciudadanos ante la decisión de crear una nueva empresa y participar con ella en la oferta de bienes y servicios, asumiendo las consecuencias de la decisión —suficiencia financiera—. La producción de bienes y servicios para la venta al mercado se realizará, por tanto, en la mayoría de los casos, en condiciones de competencia, es decir de posibilidades de elegir por parte de todos aquellos agentes que se relacionan con la empresa y, especialmente, los que compran y pagan un precio por los productos. La competencia presiona sobre la mejora continua y la innovación como respuestas para afianzar la supervivencia y obtener una recompensa acorde con los recursos empleados en la actividad. Parece, pues, realista el supuesto de la ciencia económica de analizar la razón de ser y naturaleza de las empresas desde la premisa de la eficiencia, es decir la existencia de la empresa, su naturaleza y organización interna que observamos y que el análisis positivo trata de explicar, responden al objetivo de conseguir la mejor adaptación posible a las leyes de la compe-

tencia que priman la creación de riqueza —diferencia entre valor o utilidad y coste de oportunidad.

El papel clave de los precios para coordinar —identificar desajustes entre oferta y demanda— y motivar —recompensar a quienes responden a esos desajustes en la dirección de producir más de aquello que tiene un precio más alto— en competencia a las personas que integran un colectivo social, convierte a la teoría económica en una teoría de los precios y los mercados. En esta teoría, durante mucho tiempo, apenas hubo lugar para el estudio económico de la empresa, más allá de contemplarla como un elemento, entre otros, del engranaje del mercado donde cumple una función que hace posible la formación de los precios. En efecto, los precios surgen de la intersección entre la oferta y la demanda, y para explicar la formación de precios es preciso identificar quiénes son los oferentes y los demandantes que concurren en un mercado, y, en ese interés, se encuentra un lugar para la empresa. Tal ha sido el sentido mecanicista e instrumental de la empresa en la economía de los mercados que se la ha descrito como una «caja negra», en sintonía con la absoluta indiferencia con la que la economía contemplaba su razón de ser y naturaleza.

En medio de la indiferencia de la economía como disciplina académica, la empresa gana presencia y visibilidad en la realidad social a través, sobre todo, de aumentar en dimensión y en diversidad en las formas que adopta para su funcionamiento interno. La división del trabajo se extiende al interior de las empresas de tal manera que además de las funciones o tareas propias de la producción, en las empresas se crean puestos de trabajo que tienen asignada la función de dirigir los procesos de asignación de recursos —función que supuestamente, en la lógica del mercado, le corresponde realizar al sistema de precios—. Las funciones directivas dentro de la empresa tienen complejidad suficiente como para que las personas que van a realizarlas se formen profesionalmente en ellas. Las *escuelas de negocios* se crean para dar respuesta a las necesidades formativas de los profesionales de la gestión empresarial —la Harvard Business School, una de las más prestigiosas, cumple ahora cien años.

Surge, pues, una cierta especialización entre la economía, que como disciplina académica se ocupa del estudio del funcionamiento de los mercados y la formación de precios, y las escuelas profesionales de gestión empresarial, que se ocupan de atender las demandas de formación de especialistas en puestos de dirección. La enseñanza y la investigación sobre el *management* se consolidan como ámbito para el estudio de las funciones directivas especializadas dentro de las empresas, desde la dirección de personas a la dirección general, pasando por las finanzas, el *marketing* o las operaciones. En los inicios estas enseñanzas pivotan casi exclusivamente en el estudio de casos y experiencias personales directas de los docentes. La situación cambia en los años sesenta del siglo pasado cuando el informe sobre la enseñanza de la dirección de empresas

en Estados Unidos, encargado por la Corporación Carnegie y la Fundación Ford, recomienda a las universidades impartir unas enseñanzas de gestión más fundamentales en una investigación académica rigurosa, particularmente en economía y en ciencias del comportamiento.²

Atendiendo a esta recomendación las escuelas de negocios incorporan economistas académicos, junto a profesores e investigadores de otras disciplinas científicas, tecnológicas y sociales, a sus cuadros docentes. La empresa y los procesos de gestión —*management*— se convierten en focos de creciente interés intelectual. La investigación sobre la empresa adquiere forma y cuerpo realizándose contribuciones desde muy diversas disciplinas académicas. La economía es una de estas disciplinas, de manera que la investigación económica se interesa cada vez más por la empresa en sí, sin subordinarse a los intereses del estudio del funcionamiento de los mercados. Trabajos que habían planteado retos intelectuales a los economistas académicos sobre la empresa como objeto de investigación, comienzan a recibir atención. Es el caso del artículo publicado por Ronald Coase, ya en el año 1937, sobre la naturaleza de la empresa, ignorado hasta bien entrado el siglo xx. Coase (1937) contempla la existencia de empresas y su naturaleza interna, la autoridad del empresario, como una anomalía dentro del pensamiento económico, que destaca las grandes ventajas del mercado y del sistema de precios para organizar la actividad económica: si el mercado y los precios son tan efectivos en sus funciones, se pregunta Coase, ¿por qué existen empresas donde la dirección de recursos no se realiza a partir de los precios sino según las órdenes y la autoridad del empresario?

La ortodoxia económica reconoce desde siempre las limitaciones o fallos del mercado, en determinados contextos, para lograr la concordancia entre la racionalidad individual —beneficio privado— y la racionalidad colectiva —bienestar social—. Pero, ante esas situaciones de discordancia la prescripción normativa de la economía política apunta a la intervención del Estado como forma de reconciliar los intereses en conflicto. Coase advierte que no siempre los fallos o limitaciones del mercado para dirigir —coordinar y motivar— los procesos de asignación de recursos serán resueltos a través de la intervención del Estado. Cuando sea posible —lo permitan las leyes y los costes de transacción— desde el ámbito privado surgirán instituciones —formas de dirigir la asignación de recursos que no se basan en los precios de mercado— que ayuden a superar las limitaciones del mercado sin la intervención directa del Estado. Coase contempla a la empresa como un ejemplo de institución que surge, desde el ámbito privado, cuando la coordinación de la asignación de recursos es más eficiente si se realiza a través de la mano visible del empresario que de la mano invisible del mercado. Empresa y mercado se sustituyen entre sí para realizar las funciones de organizar el intercambio, aprovechando las ventajas comparativas y sugiriendo una especialización institucional en términos de ventaja comparativa relativa.

Con el tiempo, la aportación de la economía al estudio de la empresa ha delimitado dos campos de interés que han permanecido separados hasta hace muy pocos años. Por un lado, el interés por explicar los límites o *fronteras de la empresa* y, por otro, el interés por explicar su *organización interna*. Las fronteras de la empresa se han definido horizontal y verticalmente, mientras que la organización interna distingue entre problemas de coordinación y problemas de motivación. El estudio de las *fronteras horizontales* de la empresa se ha centrado, sobre todo, en la explicación del tamaño de la empresa en términos de volumen de producción —o empleo de recursos que se requieren para ello, como por ejemplo el número de trabajadores que se emplean—. En la explicación han intervenido, principalmente, dos variables predeterminadas, la escala eficiente de producción y el tamaño del mercado. Si el mercado es suficientemente grande, la presión de la competencia forzará a las empresas a converger hacia un tamaño cercano a la escala que haga mínimo el coste unitario de producción —escala eficiente—. Las diferencias en la producción que minimiza costes unitarios —diferencias en las tecnologías de producción y grado en que están presentes los rendimientos crecientes a escala— explican la heterogeneidad de tamaños empresariales. Cuando el tamaño del mercado es reducido en relación con la escala eficiente, es de esperar que el mercado esté dominado por una empresa, configurando lo que se ha denominado monopolio natural. Desde una perspectiva dinámica, el cambio en los límites horizontales de la empresa se explica por cambios en la tecnología y en el tamaño de los mercados.

El estudio de las fronteras horizontales de la empresa forma parte del campo más amplio de la teoría neoclásica de la producción, donde la tecnología productiva se resume en una función que representa el conocimiento tecnológico más avanzado, disponible en el momento en el tiempo al que está referida, para transformar recursos en bienes o servicios de más valor o utilidad. De esta representación de la tecnología, juntamente con los precios de los recursos, se derivan las funciones de costes unitarios y las funciones de oferta a las que nos referimos anteriormente. Cuando se analiza con detenimiento, la teoría de la producción explica el tamaño de las unidades productivas —plantas de producción— pero no explica el alcance de la dirección del empresario, que define el perímetro de la empresa según Coase. La teoría no explica por qué hay unos empresarios que dirigen una sola planta de producción y otros que dirigen varias. El estudio de las fronteras de la empresa y de la organización interna —teoría de la empresa—, incorpora consideraciones de carácter contractual, donde se incluyen la disponibilidad y acceso a la información y la capacidad para procesarla, además de las meramente tecnológicas que postula la teoría de la producción. A modo de síntesis, la figura 1 resume y ordena, en el tiempo y en ámbitos temáticos, las principales contribuciones a la teoría de la empresa desde la perspectiva

² Chandler (1962, 1977) aporta documentación histórica del crecimiento de la gran empresa en Estados Unidos a principios del siglo xx y el desarrollo del manager profesional. Una visión del papel de las escuelas de negocios en la formación de directivos profesionales puede encontrarse en Kochan y Schmalensee (2003), un libro escrito con motivo del 50 aniversario de la Sloan School en MIT.

contractual en sentido amplio. Con ella se cubren el resto de materias que han reclamado el interés de la teoría económica de la empresa.

Las fronteras verticales

El estudio de las *fronteras verticales* de la empresa enlaza directamente con la observación de Coase sobre la coexistencia de los mercados y los empresarios en las funciones de coordinar la actividad económica y dirigir la asignación de recursos. Las fronteras de la empresa coinciden con el alcance de la autoridad del empresario cuando dirige la asignación de recursos, correspondiendo al mercado la coordinación entre empresas. Que el empresario dirija más o menos recursos o que dentro de la empresa se realicen más o menos actividades, que se le sustraen al mercado, depende de la eficiencia relativa de un mecanismo de coordinación u otro, la cual se ha concretado en la comparación de los costes de transacción respectivos. Una parte importante del conocimiento sobre las fronteras verticales de las empresas ha girado alrededor de los determinantes de los costes de transacción desde la perspectiva comparada: primero, empresa frente a mercado y, después, empresa, mercado y las formas intermedias de organización que incluyen a los contratos no estandarizados en las relaciones entre empresas —contratos a largo plazo, subcontratación, franquicia, alianza, empresas conjuntas, etc—. Los primeros trabajos (Arrow 1969; Williamson 1975-1985 y Klein, Crawford y Alchian 1978), se centran sobre todo en identificar los atributos que determinan las ventajas comparativas, en términos de menores costes de

transacción, de que la dirección de los recursos se realice por medio del empresario o a través del mercado. La incertidumbre y la asimetría de información entre quienes intercambian, así como la especificidad de los activos —de diversa naturaleza— invertidos en las transacciones son los atributos que, según la teoría, deben discriminar mejor a la hora de explicar la frontera vertical de las empresas; la evidencia empírica apoya estas proposiciones.

Puesto que la especificidad de los activos y las asimetrías de información, condiciones que favorecen relativamente el uso de la empresa en sustitución del mercado, concurren en un número muy elevado de transacciones económicas, la teoría de los costes de transacción (TCT) tiende a predecir un protagonismo de la empresa en la dirección de recursos superior que el que realmente se observa. Sobre esta premisa, la teoría de los derechos de propiedad (TDP), a partir de los trabajos de Grossman y Hart (1986), y Hart y Moore (1990) especialmente, aporta argumentos sobre los frenos a la expansión vertical de las empresas a partir de identificar fuentes de costes de transacción de la empresa que son ignorados por la TCT. En particular, la TDP destaca como el aspecto clave en la delimitación de las fronteras de la empresa, la distribución de la propiedad de los activos no humanos, de tal manera que para la TDP las fronteras de la empresa están delimitadas por los activos que posee en propiedad. Una implicación importante de esta visión de la empresa es que, como las personas no pueden poseerse en propiedad al quedar excluida la esclavitud, los trabajadores quedan fuera de los límites de la empresa.

Cuando la empresa amplía su presencia como mecanismo de coordinación —integrando más actividades bajo la dirección del empresario— en general está de hecho aumentando los activos no humanos bajo su propiedad, en detrimento de la propiedad de otros externos a la misma. Bajo el supuesto de que los activos no humanos son complementarios con los activos específicos que resultan de la inversión en capital humano por las personas, cuando se aumenta el control de activos no humanos por parte de una empresa, se está disminuyendo el incentivo a invertir en capital humano en aquellas personas que se ven privadas de esos activos que inicialmente poseen en propiedad. Concentrar más activos no humanos en la propiedad de una empresa aumenta los incentivos a invertir en recursos humanos específicos por parte de unas personas, pero también reduce los incentivos a invertir por parte de otras que se ven privadas de ellos. Este coste de oportunidad de ampliar los límites de la empresa con más activos justifica la predicción de la TDP por la cual, la distribución de la propiedad de los activos no humanos se dispersará en un mayor número de empresas que lo que predice la TCT.

La TCT hace hincapié en la transacción —transferencia entre unidades tecnológicamente separables— como unidad básica de análisis sobre la que se decide, en el margen, sobre si se incorpora al perímetro de una empresa o

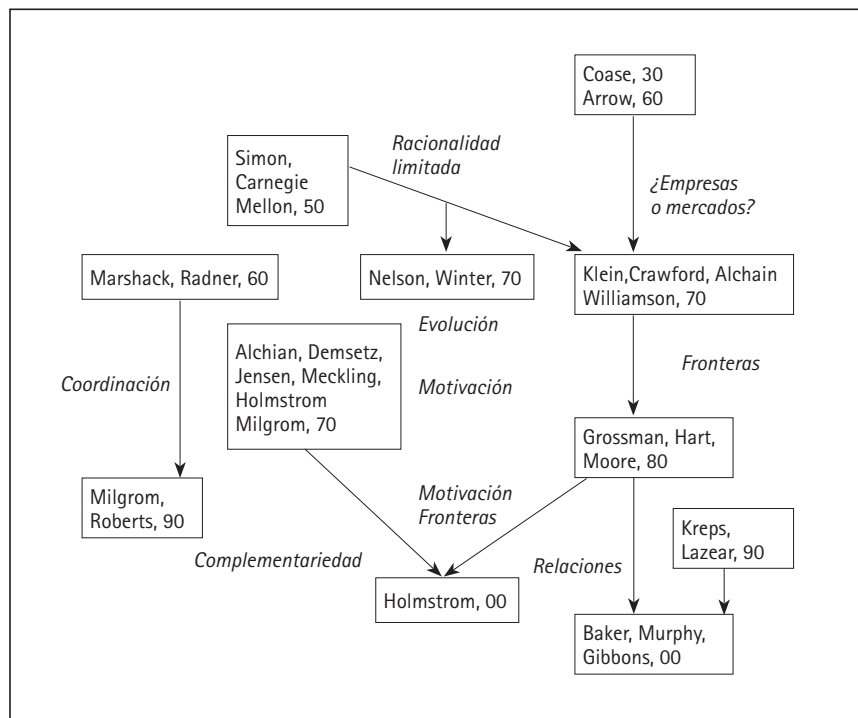


Figura 1. Aportaciones a la teoría económica de la empresa (Fuente: Salas Fumás 2007).

se deja fuera. La TDP, en cambio, destaca la decisión sobre la asignación de la propiedad de los activos no humanos, como determinante de las fronteras de la empresa. En ambos casos se aprecia un cierto distanciamiento con la perspectiva de unidad de dirección que adopta Coase para referirse a la naturaleza de la empresa. Una forma de reconciliar las diferentes perspectivas es incorporando al análisis la visión contractual. La propiedad de los activos y la jerarquía de autoridad que se atribuyen a la empresa y al empresario, tienen sentido económico en la medida en que los costes de transacción obligan a regular las transacciones por medio de contratos incompletos, es decir, contratos que delimitan el marco general por el que transcurren las relaciones entre agentes, pero donde existen muchas contingencias para los que no está previsto de antemano qué hacer cuando se producen.

La propiedad de los activos otorga la capacidad o poder para decidir sobre su uso en todo aquello que no está precontratado de antemano, mientras que la autoridad que Coase atribuye al empresario adquiere su plenitud cuando el contrato laboral es incompleto y el empresario tiene contractualmente el derecho a dirigir —mandar— a los trabajadores. Simon (1957) fue pionero en identificar la importancia económica de los contratos incompletos en las relaciones entre empresa y trabajador, aunque sin relacionar sus descubrimientos con la autoridad de origen contractual propuesta por Coase. En un mundo donde todas las relaciones entre personas se regularan a través de contratos completos —donde todo lo que puede ocurrir en cada relación está predeterminado de antemano— la propiedad de los activos no humanos y la autoridad de los empresarios no tendrían relevancia, pues no existirían derechos residuales de decisión que sí existen cuando los contratos son incompletos.³

Que los contratos tengan que ser incompletos —para evitar costes excesivos de transacción— junto con la repetición de las relaciones entre agentes, abre el camino a otro tipo de contratos, los implícitos, cuya viabilidad y eficacia condiciona las decisiones sobre las fronteras de la empresa. En efecto, los contratos implícitos son aquellos en los que los acuerdos entre las partes se sustentan por la expectativa de buena fe en la conducta ajena y la confianza mutua. La confianza emerge más fácilmente en relaciones interpersonales donde el poder de decisión está repartido entre los agentes implicados, que en situaciones donde el poder de decisión está muy polarizado. Para aprovechar las ventajas, en términos de bajos costes de transacción, de los contratos implícitos, la decisión sobre la frontera de la empresa, en términos de asignación de la propiedad de activos no humanos empleados en la producción, primará una cierta dispersión de la propiedad como una forma de equilibrar el poder y fomentar la confianza como soporte de las relaciones entre agentes en las sucesivas etapas de la cadena productiva (Baker, Murphy y Gibbons 2001).

Organización interna

La investigación económica ha separado el estudio de las fronteras o límites de la empresa del estudio de lo que ocurre dentro de la empresa, para unas fronteras dadas. En buena parte, el estudio interno de la empresa se ha concentrado, más o menos explícitamente, en analizar la forma en que el empresario dirige la asignación de recursos, así como la razón de ser del propio empresario. La organización interna de la empresa —el orden que resulta de la división interna del trabajo, el intercambio de información y el reparto del poder para la toma de decisiones— se consigue por medio de la coordinación —determinar qué debe hacer cada uno— y la motivación —tener el interés por hacerlo— de las personas implicadas. La diferencia con respecto al mercado es que en la empresa la forma de abordar la coordinación y motivación está mediatizada por la posibilidad de intervención del empresario.

La investigación económica comienza separando el estudio de los problemas de coordinación del de los problemas de motivación, aunque los dos han girado de forma significativa alrededor de un concepto muy arraigado en la realidad de las empresas y su gestión como es el concepto de *equipo*. Para el estudio de la coordinación se postula el concepto de *organización de equipo*, entendida como acción colectiva donde todas las personas que colaboran en ella comparten el mismo objetivo que, además, representa el interés del grupo, por ejemplo hacer máxima la riqueza conjunta. Aunque cada persona persigue en sus actuaciones el bien general, lo que debe decidir o hacer para conseguir la máxima eficiencia en el resultado depende de la información y las acciones de los demás —interdependencias—. Coordinar las acciones individuales significa, en este contexto, influir en las decisiones individuales —principalmente a través de intercambiar información— para compatibilizarlas dentro de unas interdependencias dadas, o bien alterar las interdependencias para influir así en las propias necesidades de coordinación (Marshack y Radner 1972; Milgrom y Roberts 1995). En organizaciones de equipo la coordinación se convierte en un problema relevante porque existe un coste no despreciable en la producción y transmisión de información. Por ello, las soluciones al problema de coordinación estudiadas en la literatura, y también las puestas en marcha por las empresas, han tenido mucho que ver con los avances en la tecnología de la información.

Por otra parte, la producción o *tecnología de equipo* se refiere a la complementariedad entre recursos propiedad de personas diferentes que colaboran en la producción, de manera que el aprovechamiento conjunto de la tecnología crea más riqueza potencial que el aprovechamiento por separado. Alchian y Demsetz (1972), que sitúan esta característica de la tecnología en el origen de la empresa tal como la conocemos, analizan el funcionamiento colectivo de la producción en equipo bajo el supuesto opuesto a la organización de equipo, es decir, las personas que aportan los recursos complementarios a la produc-

3

El referente empírico de la teoría sobre las fronteras de la empresa lo constituyen las decisiones empresariales de fabricar, comprar o aliarse con terceros. En los últimos años, a estas decisiones se les han añadido términos como *outsourcing* y *offshoring*, para remarcar la reubicación de las actividades de la cadena de valor de las empresas en los mercados globales. La teoría de las fronteras de la empresa ha entrado de lleno en el estudio de la empresa multinacional (Helpman 2006).

ción lo hacen con la expectativa de obtener una máxima recompensa neta individual, sin preocuparse por el interés colectivo. Tecnología más organización de equipo dan lugar exclusivamente a problemas de coordinación; tecnología de equipo más superposición de intereses individuales —también llamada coalición— dan lugar a problemas de coordinación más otros problemas de motivación, que son los que han centrado el mayor interés académico.

La tecnología de equipo impide dividir la producción conjunta como suma de las producciones individuales de quienes colaboran en esa producción. Las contrapartidas individuales que inducen la participación en la acción colectiva y motivan la aportación de recursos por parte de cada agente implicado, sólo pueden definirse en función de la producción conjunta y de las aportaciones de recursos, si son observables. En principio, compensar la participación en la acción colectiva con una participación en el *output* conjunto tiene ventajas en cuanto que sólo necesita medir una variable; sin embargo tiene inconvenientes conocidos en forma de conductas de polizón (Holmstrom 1982). La alternativa de medir las contribuciones de recursos —cantidad y calidad— exige especialización en esa tarea y dar respuesta a la pregunta de cómo se asegura la eficiencia de quien monitoriza el proceso. Alchian y Demsetz responden con la propuesta de una solución organizativa donde el monitor contrata de forma bilateral con cada participante en la acción colectiva, pacta una contrapartida acorde con la cantidad de recurso que se aporta, adquiere el derecho a supervisar su actuación y dirigir su trabajo y retiene, como retribución propia, la diferencia entre lo que se produce y lo que se ha comprometido a pagar. En suma, la teoría de la empresa da sentido económico a la empresa capitalista tal como la conocemos, donde el empresario centraliza los contratos y asume la competencia supervisora, además de la coordinación, a cambio del beneficio residual.

Sobre la base del paradigma de producción en equipo, carácter bilateral de los contratos y renta residual —beneficio— del empresario, han prosperado sucesivas aportaciones a la teoría de la empresa, bajo la caracterización de la misma como nexo de unión entre contratos. Los desarrollos desde los años setenta han tenido en cuenta las limitaciones del supervisor para medir con precisión la cantidad y calidad de los recursos, así como la desigual asignación de los riesgos que lleva consigo la retribución en forma de beneficio cuando el resultado de la acción colectiva depende de factores aleatorios además de los recursos empleados y la tecnología disponible. Surgen así aportaciones al diseño organizativo de la empresa como: a) determinar el número de niveles jerárquicos de supervisión y tramo de control (Calvo y Wellisz 1978; Rosen 1982); b) asignar de forma eficiente los riesgos, por ejemplo, a través de separar en colectivos diferentes las funciones de dirección de recursos, coordinar y motivar, y las funciones de asunción de riesgos, lo cual da lugar a la empresa capitalista compleja o corporación (Jensen y Meckling 1976);

c) determinar la ubicación óptima del poder de decisión (centralización *versus* descentralización: Aghion y Tirole 1997; Alonso, Dessein y Matouschek 2008); d) diseñar sistemas de incentivos complejos para estimular el esfuerzo no observable por el supervisor (teoría de la agencia: Holmstrom 1979, 1982; Holmstrom y Milgrom 1987, 1991).⁴

Con todos estos desarrollos en la teoría sobre fronteras y organización interna de la visión de la empresa, se distancia de la función de producción para convertirse en una estructura que gobierna el proceso de asignación de recursos procesando información, asignando poder de decisión, evaluando el desempeño y otorgando recompensas.

La empresa como mini economía y como comunidad de personas

La teoría económica de la empresa ha utilizado casi indistintamente los términos de empresa y empresario, aunque de forma inconsciente. La lectura atenta de las obras de Coase y de Alchian y Demsetz pone de manifiesto que lo que se explica en ellas es la existencia del empresario con unas funciones concretas que realiza dentro de un marco general de especialización y división del trabajo: al empresario coasiano le corresponde dirigir los recursos con coordinación, dando órdenes y aprovechando su posición central en la red de contratos con otros propietarios de recursos, allí donde el mercado y los precios afrontan des-economías de escala. En cuanto al empresario supervisor de Alchian y Demsetz, su función es necesaria para medir las aportaciones en cantidad y calidad de los recursos que nutren la producción con tecnología de equipo. Las visiones contractuales de la empresa como nexo común de contratos que se han inspirado en estos autores, no distinguen si el nexo de unión es la persona del empresario o algo sin precisar que llaman empresa. Podría decirse, por tanto, con razón, que muchas de las teorías de la empresa que se han propuesto desde la economía son, en realidad, teorías del empresario. El empresario director que emerge de esta literatura complementa la visión schumpeteriana de este agente social como protagonista de la innovación y de la destrucción creativa.

Para deslindar la empresa del empresario la mejor fórmula es contemplar la empresa como persona jurídica o entidad legal, reconocida por el derecho para participar en contratos y para poseer activos en propiedad. El nexo común de contratos con que se identifica a las empresas es, en la mayoría de los casos, una persona jurídica que contrata bilateralmente con los diferentes agentes con los que se relaciona. Para que el empresario tenga capacidad para coordinar y motivar a las personas dentro de la empresa, es preciso que el contrato de la empresa con estas personas contemple la posibilidad de que un tercero, también vinculado contractualmente con la empresa, realice esas funciones. En la TDP, donde las fronteras de la empresa se relacionan con los activos no humanos que posee en propiedad, es preciso explicar por qué esa propiedad recae en el ente jurí-

4 Cuando se revisa esta literatura se comprueba que más que configurar una teoría de la empresa define un campo de estudio, la economía de la información. Muchos de los problemas de transacción en condiciones de información asimétrica que se analizan con los modelos de selección adversa o riesgo moral, no son específicos de las empresas sino que ocurren también en el dominio de los mercados. Por tanto, la teoría de la agencia no es una teoría de la empresa sino un marco teórico para estudiar problemas de riesgo moral, algunos de los cuales forman parte de los problemas de organización interna que afrontan las empresas.

dico de la empresa y no en la persona física del empresario. En suma, la teoría económica de la empresa no estará completa hasta que se explique por qué surge el ente legal de la empresa diferente de la persona física del empresario.

En la literatura existen varias respuestas posibles, todas ellas relacionadas de algún modo con el deseo de economizar en costes de transacción:

I. La persona jurídica de la empresa no está afectada por la limitación en el horizonte temporal de carácter finito, que sí afecta a las personas físicas. Un horizonte de más largo plazo y no finito es relevante para la viabilidad de las relaciones basadas en la reciprocidad que sostiene la confianza mutua —contratos implícitos— y da valor económico a una buena reputación (Kreps 1990).

II. La entidad legal, complementada con la diversidad de formas jurídicas entre las que el derecho permite elegir al constituir una empresa, proporciona posibilidades de gestionar los riesgos, transferir la propiedad, especializar entre dirigir los recursos de la empresa y financiar sus activos, que no serían posibles si las personas no pudieran diferenciar entre el patrimonio personal y el patrimonio propiedad de las empresas. La entidad intelectual y técnica de las cuestiones que plantea la elección de la forma jurídica de la empresa, y el dar respuesta a las opciones abiertas sobre cómo asignar competencias de gestión y competencias de control, han dado lugar a un ámbito de estudio sobre la empresa capitalista moderna de gran relevancia como el del *gobierno corporativo*, donde confluyen el derecho y la economía.⁵

III. Con la concentración de la propiedad de los activos no humanos en la persona jurídica de la empresa, frente a la alternativa de que la propiedad de los activos se reparta entre las diferentes personas que se vinculan a través de ella, la dirección de la empresa encuentra formas eficientes de coordinar y motivar a los trabajadores en contextos de información asimétrica, que no serían factibles en el caso de que la propiedad de los activos se repartiera entre todos los trabajadores (Holmstrom 1999). Con esta explicación de la persona jurídica de la empresa como propietaria de activos, Holmstrom integra en un único problema de diseño organizativo óptimo las decisiones que tienen que ver con las fronteras de la empresa —asignar la propiedad de los activos no humanos—, y las decisiones para coordinar y motivar el trabajo dentro de la empresa, que son competencia de la organización interna. Con la visión integradora de la teoría de la empresa, Holmstrom llega a definir a la empresa como una *mini economía*, donde los responsables de la dirección afrontan la solución de ineficiencias originadas en problemas de información asimétrica, y efectos externos de naturaleza similar a los que afronta la autoridad el Estado para el conjunto de la sociedad. Aunque con la diferencia importante de que la empresa está rodeada de mercados, que ofrecen opciones de salida que limitan los posibles excesos de poder y que derivan de la gran concentración de activos que se pueden acumular.

La empresa como ente legal propietaria de activos sobre cuyo acceso y condiciones de uso deciden quienes las dirigen, se convierte en una palanca de poder que incide sobre la conducta de las personas que combinan su trabajo y sus conocimientos con dichos activos. Por esta razón, aunque formalmente es verdad que la empresa no posee en propiedad al capital humano incorporado en las personas y, por tanto, éstas quedan fuera de su perímetro, el funcionamiento de la empresa en su conjunto se entiende mejor si los trabajadores se contemplan como parte de ella. Una teoría de la empresa que incorpore a las personas que trabajan combinando su esfuerzo y conocimientos con los activos de su propiedad, supone un acercamiento importante a las aproximaciones al estudio de las empresas desde otras disciplinas sociales como la psicología y la sociología, que desde un principio contemplan la empresa como una comunidad de personas, minimizando la relevancia del resto de los activos, lo opuesto a lo que sucede con la economía. Por otra parte, el vínculo entre la economía y el resto de disciplinas sociales para el estudio de la empresa se estrecha a partir de visiones más relajadas del concepto de racionalidad, que ha primado entre los economistas y otros desarrollos recientes.

La empresa como comunidad de personas

El estudio económico de la empresa se realiza bajo la premisa del comportamiento humano, que se resume en *racionalidad*: las personas conocen sus preferencias y su comportamiento es coherente con ellas. La racionalidad permite a la investigación académica simular las consecuencias privadas y sociales de determinadas conductas y restricciones, recomendando correcciones o ajustes en función de los resultados previstos. Sin embargo, la racionalidad económica ha recibido críticas por el poco realismo de los supuestos que contempla y por los aspectos del comportamiento humano que deja sin explicar. Williamson (1975), en el libro que recupera la comparación institucional entre mercado y empresa —jerarquía— que había planteado cuarenta años antes Coase, incorpora a la teoría de la empresa la crítica a la hipótesis de racionalidad que habían adelantado Herbert Simon y sus colegas de Carnegie Mellon, al advertir de que los supuestos sobre capacidad de almacenar y procesar información que están implícitos en la maximización de la utilidad que dispone la economía, son poco realistas teniendo en cuenta la fisiología del cerebro humano.

Las críticas de Simon y sus colaboradores llevan a proponer un supuesto alternativo a la racionalidad absoluta bajo la que se estudia la empresa y el mercado en la economía, conocido como *racionalidad limitada*: las personas son intencionadamente racionales pero su comportamiento se ve afectado por limitaciones en la capacidad de almacenar y procesar información tan relevantes para explicar la realidad como las restricciones que provienen de la tecnología. Bajo estos supuestos, la explicación del

5

El gobierno corporativo incorpora un tema más general que es el de la decisión sobre la propiedad de la empresa (Hansmann 1996). Tirole (2001) ofrece una visión del problema de gobierno de la empresa que se ajusta a la perspectiva contractual que se adopta en este texto.

comportamiento humano incorpora supuestos de decisión heurística frente a la optimización que se presume desde la racionalidad sin limitaciones. La evolución y la adaptación en procesos de tránsito de unos equilibrios a otros son estados permanentes del sistema, que no pueden ignorarse como hace la economía neoclásica, que sólo compara situaciones de equilibrio (Nelson y Winter 1982).

Experimentos de laboratorio y la observación de la realidad proporcionan evidencias sobre comportamientos humanos que no encajan con los supuestos de consistencia y transitividad de preferencias propios de la racionalidad más convencional, y han impulsado un campo de especialización en *economía del comportamiento* —*behavioral economics*—, donde destacan la *teoría prospectiva* o del punto de referencia, Kahneman y Tversky (1979), que pone en cuestión la teoría de la utilidad esperada sobre la que se analiza buena parte del comportamiento en situaciones de riesgo en el estudio de la empresa y los mercados; o la teoría de las *preferencias sociales* (Guth et al. 1982), que cuestiona el supuesto clásico de que las personas sólo tienen en cuenta sus propios pagos cuando eligen entre alternativas (véase Camerer, Loewenstein y Rabin 2004, para una revisión de esta literatura). La economía del comportamiento supone un acercamiento importante entre la economía y la psicología; uno de los últimos pasos en el proceso es la importación, desde la psicología, del concepto de «felicidad» como alternativa a la utilidad que tradicionalmente utiliza la economía, para expresar las preferencias de las personas (Frey 2008). La economía clásica renuncia a la introspección para explicar cómo se forman las preferencias de las personas y opta, en cambio, por las preferencias reveladas como un medio para inferir la utilidad de las alternativas planteadas apelando al supuesto de racionalidad —consistencia y coherencia entre preferencias y conducta—. La investigación sobre la felicidad, en cambio, adopta formas de medir la utilidad que se han desarrollado en la psicología y, dentro de ella, en la neurociencia, con un objetivo claramente introspectivo. Los resultados experimentales de estas investigaciones muestran que las personas no sólo valoran bienes y servicios tangibles a los que se tiene acceso a través de la renta monetaria, sino que en su utilidad intervienen también aspectos menos tangibles como las condiciones sociales, relaciones, la capacidad de decidir y la posibilidad de desarrollar sus propias competencias. Surge así la *utilidad asociada a los procesos*, en contraste con la utilidad basada sólo en los resultados, que predomina en la economía más ortodoxa.

La economía del comportamiento está modificando la forma de analizar el funcionamiento de los mercados —Frey se refiere a la felicidad como un concepto que revolucionará la ciencia económica— y también ayuda a explicar ciertas particularidades del funcionamiento de las empresas, que a veces se contemplan como anomalías del modelo neoclásico. Por ejemplo, una de las regularidades

que se observan en las empresas es la estabilidad de las relaciones entre ellas y los trabajadores, con contratos a largo plazo y mercados internos de trabajo —la promoción interna es el mecanismo dominante para cubrir vacantes en puestos de trabajo, en lugar de acudir a la contratación externa—. La estabilidad de las relaciones obedece, en parte, a la imposibilidad de adquirir en el mercado conocimiento y saber específico que sólo se acumula a través del aprendizaje de rutinas que surgen del ajuste mutuo y de la evolución adaptada a las condiciones de un contexto determinado. En este sentido las fronteras de la empresa se han explicado a partir de las necesidades de proteger e intercambiar un conocimiento específico y valioso, que a la vez aporta ventajas competitivas en el mercado (Teece 1986; Kogut y Zander 1992).

Una consecuencia de la estabilidad, persistencia y carácter rutinario de muchas de las relaciones entre las personas que colaboran alrededor del nexo común de una empresa es que las personas pueden acabar formando una comunidad social, donde las relaciones interpersonales pueden estar condicionadas por sentimientos y sentido de la reciprocidad, como ocurre en cualquier grupo humano estable (Gintis et al. 2005). El concepto de *cultura* como forma generalmente aceptada y espontánea de conducir las relaciones interpersonales en una comunidad social, aparece de forma natural en el funcionamiento de la empresa. La cultura, concepto de raíces sociológicas, permite soluciones a los problemas de coordinación y motivación que son difíciles de explicar desde una racionalidad ajena al contexto de normas y valores del que forman parte las personas —por ejemplo, la aparición de conductas de cooperación en entornos de dilema de prisionero donde la racionalidad clásica predice un comportamiento no cooperativo—, aspecto que no ha pasado desapercibido a economistas ortodoxos (Kreps 1990; Kandel y Lazear 1992).

Estos trabajos teóricos amplían el modelo convencional de preferencias y racionalidad para dar cabida a regularidades empíricas, que aparecen en un momento histórico en que el modelo de empresa dominante en Japón —organización por procesos y en equipos con gran autonomía orientada a los clientes— pone en cuestión los modelos de empresa que se habían forjado bajo el referente empírico del modelo de empresa dominante en Estados Unidos —la empresa jerárquica—, sobre todo al comprobarse el éxito comercial de las empresas japonesas en los mercados donde compiten con las empresas americanas. Las investigaciones sobre la felicidad suponen un nuevo reto para la teoría de la empresa y para las prácticas de gestión que se implanten en ella, en cuanto que las preferencias de las personas sobre cómo se consiguen los resultados, además de los propios resultados, obliga a valorar la organización interna —diseño de puestos de trabajo, autonomía de las personas en ellos, participación en las decisiones, mecanismos de socialización— como fin en sí mismo y no sólo como un medio para conseguir mejores resultados.

Conclusión

El conocimiento académico sobre la empresa resulta imprescindible para comprender el funcionamiento del conjunto de la economía, en cuanto que lo que ocurre dentro de las empresas es cuantitativa y cualitativamente tan importante como lo que ocurre entre las empresas (Simon 1991). Resulta difícil exponer una teoría de la empresa, aunque sea dentro de una sola disciplina académica como la economía, porque los propios desarrollos intelectuales alrededor de la misma no son del todo precisos en la acotación del concepto. En este sentido, la empresa aparece asociada de forma más o menos explícita con una unidad técnica —planta— de producción, con la función del empresario o persona en quien delega la dirección de recursos, con un ente legal creado por el derecho o con una comunidad de personas. A veces la teoría de la empresa responde a la pregunta sobre los determinantes de sus límites o fronteras, y otras veces a preguntas que tienen que ver con la solución a problemas de motivación y coordinación internos.

En todo caso, desde el análisis económico, la empresa cumple la función de producir bienes y servicios para el mercado en condiciones de competencia y suficiencia financiera, pudiendo adoptar para ello una de las múltiples formas jurídicas contempladas por el derecho. Un aspecto que diferencia a la empresa dentro del mercado del que forma parte es su condición de nexo de contratos, siendo la condición de persona jurídica la que permite a la empresa ocupar el lugar del empresario en el nexo común. La interposición de empresas evita los contratos multilaterales entre inversores, trabajadores, clientes que serían necesarios en una solución de mercado, con el consiguiente ahorro en costes de transacción; que el nexo sea una persona jurídica ofrece posibilidades para la acumulación de activos y la gestión de riesgos, así como para la gestión interna en condiciones de información asimétrica y efectos externos, que no son alcanzables para personas físicas ocupando el mismo lugar. Aunque cada forma jurídica impone ciertas restricciones a la forma de resolver los problemas de coordinación y motivación —en aras a reducir costes de transacción en el agregado— también deja suficientes grados de libertad como para que, en cada caso, se adopten las soluciones a estos problemas más acordes con las características de las transacciones en que se ven envueltos los agentes. Por otra parte, la forma dominante de empresa cambia en el tiempo y, entre países de nivel de desarrollo económico similar, en

el mismo periodo de tiempo, lo cual no debe hacer olvidar que la empresa es una invención humana y, como tal, susceptible de modificación y transformación atendiendo a condiciones tecnológicas, por ejemplo, los desarrollos en las tecnologías de la información, e institucionales del entorno —sistemas legales—.⁶ La teoría de la empresa aspira a identificar problemas básicos de coordinación y motivación, que tienen permanencia estructural pero que permiten diferentes soluciones en condiciones de entorno también diferentes.

El conocimiento sobre la empresa como respuesta institucional a los problemas de intercambio y colaboración que resultan de la división del trabajo, no debe confundirse con el conocimiento sobre la dirección de empresas —*management*— que se transmite a quienes ocupan o quieren ocupar puestos de dirección empresarial. Parece recomendable que el conocimiento positivo sobre la empresa, que proporciona la teoría, forme parte del conocimiento normativo sobre cómo dirigir una empresa, que se imparte en las escuelas de negocios, aunque en la realidad no es así. Una forma de explicar este distanciamiento entre el conocimiento normativo y el positivo es que la teoría de la empresa presupone una racionalidad absoluta en el comportamiento de los agentes, y su interés se centra exclusivamente en conocer las implicaciones para el bienestar colectivo de esa racionalidad individual. Esta explicación falla cuando las decisiones y las conductas de quienes dirigen las empresas son difíciles de reconciliar con la racionalidad sobre la que se construye la teoría. Los avances en la economía del comportamiento y la incorporación de la felicidad a la introspección de las preferencias y de la utilidad, servirán para acercar el conocimiento positivo y el conocimiento normativo sobre la empresa. Por este camino la empresa puede perder su condición de instrumento o medio para obtener los mejores resultados económicos posibles —mayor renta y consumo—, que ha tenido en el análisis económico convencional, para pasar a ser valorada en función de las soluciones concretas que adopta ante los problemas de coordinación y motivación que surgen en su seno. Los medios y los resultados deberán valorarse conjuntamente, aunque es posible que socialmente se primen unos sobre los otros. Las referencias a la ética y a la responsabilidad social de las empresas en los últimos años, bien pueden indicarnos que la sociedad está mostrando preferencias por el cómo más que por los resultados cuando se valora la *performance* de las empresas.

6

Con el desarrollo de las tecnologías de la información y las comunicaciones (TIC) aparecen nuevos modelos de empresa y de negocio empresarial, Microsoft, Intel, Cisco, que sustituyen como referentes empíricos a tener en cuenta a Ford y GM, hasta los años ochenta, y a Toyota en la última parte del siglo xx. El nuevo modelo empresarial hace referencia a la virtualidad y a la estructura de red que adopta la organización empresarial, y rompe con la tradición de relaciones duraderas —empleo vitalicio, contratos a largo plazo con proveedores, fidelización de clientes— que domina en la empresa del pasado. La visión de la empresa como comunidad humana, si tiene razón de ser en el nuevo escenario, deberá reinventarse (Castells 1996).

Este texto se basa en gran parte en el libro del autor, *El siglo de la Empresa*, publicado por la Fundación BBVA.

Bibliografía

- Aighon, P. y J. Tirole. «Formal and real authority in organizations». *Journal of Political Economy*, 105 (1997): 1-29.
- Alchian, A. y H. Demsetz. «Production, information and economic organization». *American Economic Review*, 62 (1972): 777-795.
- Alonso, R., W. Dessein y N. Matouschek. «When does coordination require centralization?». *American Economic Review*, 98 (2008): 145-179.
- Arrow, K. «The organization of economic activity. Issues pertinent to the choice of market versus non market resource allocation». En US Joint Economic Committee, *The analysis and evaluation of public expenditures. The PBS system*, 1969.
- Baker, G., R. Gibbons y K. Murphy. «Relational contracts and the theory of the firm». *Quarterly Journal of Economics*, 117 (2001): 39-83.
- Calvo, G., S. Wellisz. «Supervision, loss of control and the optimal size of the firm». *Journal of Political Economy*, 87 (1978): 943-952.
- Camerer, C., G. Loewenstein, M. Rabin. *Advances in Behavioral Economics*. Princeton: Princeton University Press, 2004.
- Castells, M. *The Rise of the Network Society. Vol. 1: The Information Age: Economy, Society and Culture*. Berkeley: University of California Press, 1996.
- Chandler, A. *Strategy and Structure*. Cambridge: MIT Press, 1962.
- *The Visible Hand*. Cambridge: Belknap Press, 1977.
- Coase, R. «The nature of the firm». *Economica*, 4 (1937): 386-405.
- DiMaggio, P. *The Twenty-First-Century Firm*. Princeton: Princeton University Press, 2001.
- Frey, B. *Happiness: A Revolution in Economics*. Cambridge: MIT Press, 2008.
- Gintis, H., S. Bowles, R. Boyd y E. Fehr. *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*. Cambridge y Londres: MIT Press, 2005.
- Grossman, S., O. Hart. «The costs and benefits of ownership: A theory of lateral and vertical integration». *Journal of Political Economy*, 94 (1986): 691-719.
- Guth, W., R. Schmittberger, B. Schwarze. «An experimental analysis of ultimatum bargaining». *Journal of Economic Behavior and Organization*, 3, 1982: 367-388.
- Hansmann, H. *The Ownership of Enterprise*. Cambridge: Belknap Press, 1996.
- Hart, O., J. Moore. «Property rights and the theory of the firm». *Journal of Political Economy*, 98 (1990): 1.119-1.158.
- Hayek, F. «The use of knowledge in society». *American Economic Review*, 35 (1945): 519-530.
- Helpman, E. «Trade, FDI and the organization of firms». *Journal of Economic Literature*, 3 (2006): 589-630.
- Holmstrom, B. «Moral hazard and observability». *Bell Journal of Economics*, 10, 1979: 74-91.
- «Moral hazard in teams». *Bell Journal of Economics*, 13 (1982): 324-340.
- «The firm as a subeconomy». *Journal of Law Economics and Organization*, 1999: 74-102.
- y P. Milgrom. «Multitask principal-agent analysis: Incentive contracts, asset ownership and job design». *Journal of Law, Economics and Organization*, 7 (1991): 24-52.
- y P. Milgrom. «Aggregation and linearity in the provision of intertemporal incentives». *Econometrica*, 55 (1987): 308-328.
- Jensen, M. y W. Meckling. «Theory of the firm: managerial behaviour, agency costs and ownership structure». *Journal of Financial Economics* 3 (1976): 305-360.
- Kahneman, D. y A. Tversky. «Prospect theory: An analysis of decision making under risk». *Econometrica*, 47 (1979): 263-291.
- Kandel, E. y E. Lazear. «Peer pressure and partnerships». *Journal of Political Economy*, 100 (1992): 801-817.
- Klein, B., R. Crawford y A. Alchian. «Vertical integration, appropriable rents and the competitive contracting process». *Journal of Law and Economics*, 21 (1978): 297-326.
- Kochan, T. y R. Schmalensee. *Management: Inventing and delivering its future*. Cambridge: MIT Press, 2003.
- Kogut, B. y U. Zander. «Knowledge of the firm, combinative capabilities and the replication of technology». *Organization Science*, 3 (1992): 383-379.
- Kreps, D. «Corporate Culture», en J. Alt, K. Shepsle eds. *Perspectives on Positive Political Economy*. Cambridge: Cambridge Uni. Press, 1990.
- Malone, Th., R. Laubacher y M. Scott Morton (eds.). *Inventing the Organizations of the 21st Century*. Cambridge: MIT Press, 2003.
- Marschak, J. y R. Radner. *Economic Theory of Teams*. New Haven: Yale University Press, 1972.
- Milgrom, J. y J. Roberts. «Complementarities and fit: Strategy, structure and organizational change in manufacturing». *Journal of Accounting and Economics*, 19 (1995): 179-208.
- Milgrom, P. y J. Roberts. *Economics, Organization and Management*, Nueva Jersey: Prentice Hall, 1992.
- Nelson, R. y S. Winter. *An Evolutionary Theory of Economic Change*. Cambridge: Belknap Press, 1982.
- Roberts, J. *The Modern Firm*. Oxford: Oxford University Press, 2004.
- Rosen, S. «Authority, control and the distribution of earnings». *Bell Journal of Economics*, 13 (1982): 311-323.
- Simon, H. «A formal theory of employment relationship». *Econometrica*, 19 (1951): 293-305.
- «Organizations and markets». *Journal of Economic Perspectives*, 5 (1991): 25-44.
- Teece, D. «Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy». *Research Policy*, 15 (1986): 285-305.
- Tirole, J. «Corporate governance». *Econometrica*, 69 (2001): 1-35.
- Williamson, O. *Markets and Hierarchies*. Nueva York: Free Press, 1975.
- *The Economic Institutions of Capitalism*. Nueva York: Free Press, 1985.

fronteras de la ciencia económica

JOAN ESTEBAN

Introducción

Este artículo trata sobre las diversas tendencias en el ámbito de la teoría económica a lo largo de las últimas décadas, y es completamente subjetivo y parcial. Con él no aspiro a presentar una visión global de la investigación actual en el terreno de la ciencia económica, más bien he preferido centrarme en algunos estudios recientes en los que se ha intentado ampliar los rígidos supuestos sobre los que se formuló la *teoría del equilibrio general* (EG) durante la segunda mitad del siglo xx. Los intentos por ampliar la descripción del comportamiento de individuos, empresas, gobiernos y la sociedad en general han tenido el efecto secundario de incrementar de manera significativa el carácter interdisciplinar de la investigación en ciencia económica. Estamos asistiendo a un solapamiento significativo con ciencia política y sociología, por supuesto, pero también con psicología, biología y neurociencia.

A pesar de que el tema de este ensayo es ya de por sí específico, me veré obligado a tratarlo de forma más superficial de lo que me gustaría. Además, mi elección tiene el inconveniente de que obliga a pasar por alto áreas importantes y dinámicas de la economía como son la macroeconomía, las finanzas, el comercio y el desarrollo, por mencionar sólo algunas.

La estructura de este ensayo es la siguiente: en el apartado que sigue haré una breve exposición del modelo de EG,

que sin duda constituye el paradigma central de la ciencia económica. En el apartado tercero se exponen los principales puntos en los que la investigación se ha apartado del modelo estándar de EG. A continuación analizaré con mayor detenimiento las aportaciones recientes del enfoque *conductista* del proceso de toma de decisiones de los individuos. La investigación del comportamiento, que se ha inspirado en la psicología y que se basa sobre todo en experimentos controlados, intenta documentar cuidadosamente las pautas de comportamiento de los individuos que se alejan de las previstas por el modelo tradicional de comportamiento racional. En el apartado cuarto se hace una exposición detallada de todos los elementos que intervienen en la toma de decisiones, de forma que podamos explicar de manera más estructurada los diferentes resultados y cuáles de ellos exactamente se ponen en tela de juicio. En el quinto se presentan brevemente las contribuciones en el ámbito de la *economía conductista*. Por último, el apartado 6 hace balance de las investigaciones presentadas, evalúa sus aportaciones y saca conclusiones para la investigación venidera.

El equilibrio general y la economía del bienestar

La teoría moderna del EG comenzó en la década de 1950. El artículo de 1954 de K. Arrow y G. Debreu, ambos premio Nobel de Economía, acerca de la existencia de un equilibrio competitivo, y el libro de Debreu publicado en 1959,

Teoría del valor, marcaron el comienzo de cuatro décadas en las que se llevaron a cabo esfuerzos extremadamente fructíferos para entender el funcionamiento de los mercados competitivos. Se puede considerar que las obras de Arrow y Hann, *General Competitive Analysis* (1971); de W. Hildenbrand, *Core and Equilibria of a Large Economy* (1974), y de A. Mas-Colell, *The Theory of General Equilibrium: A Differentiable Approach* (1990) constituyen los hitos más significativos de este empeño.

La teoría de EG es un modelo muy complejo pero esquemático que trata de reproducir el papel coordinador que desempeñan los mercados en una sociedad fraccionada e individualizada. Constituye la prueba rigurosa de la afirmación de Adam Smith de que basta con *una mano invisible* para hacer que las decisiones que toman individuos y empresas entre los que no existe coordinación alguna sean compatibles entre sí.¹ En este sentido, era crucial para este modelo que se eliminase todo motivo de coordinación explícita o tácita, así como que los individuos pudiesen tomar decisiones en función de los demás. Con todos los participantes persiguiendo tan sólo su propio interés —en sentido estricto— y limitando su interacción a ofrecer y demandar mercancías a través de los mercados, el resultado no sería caótico, sino ordenado y eficaz (en un sentido que se explicará más adelante).

Recordemos algunos de los supuestos necesarios para obtener estos resultados matemáticos. Los agentes que participan son un número arbitrario de individuos y empresas, y todos ellos pueden comprar y vender bienes (su trabajo, por ejemplo) en los mercados y a precios vigentes. Y, lo que es aún más importante, se supone que todos los participantes se comportan *de manera competitiva*, esto es, aceptan los precios como dados.² Además, los individuos poseen todas las acciones de todas las empresas, cuya distribución entre la población es arbitraria.

En un principio cada individuo tiene una serie de bienes (que puede simplemente ser trabajo) y acciones. Al comprar y vender, pueden obtener otras nuevas mercancías (por ejemplo, vende ocho horas de trabajo y compra pan y mantequilla). Las cantidades compradas y vendidas por cada individuo deben estar dentro de su presupuesto, cuyo valor monetario vendrá determinado por los precios vigentes. Todos ellos tienen un orden de preferencia con respecto a las diferentes mercancías, esto es, pueden ordenar cualquier par de cestas de mercancías según su deseabilidad. Además de completo, se supone que este orden es transitivo, reflexivo y satisface una condición de convexidad débil.³ Lo que es más importante es que estas preferencias dependen sólo del propio consumo, con lo que se elimina todo tipo de altruismo. También se entiende que los individuos actúan de manera racional, esto es, dentro de los bienes a su alcance —el valor de las compras no puede exceder al de las ventas— eligen el que es más deseable según su orden de preferencias. Por consiguiente, para todo vector dado de precios de mercado cada consu-

midor tiene, asimismo, un vector bien definido de oferta y demanda de bienes.

Las empresas compran a través del mercado bienes y trabajo que suministran otras empresas y otros consumidores individuales (*inputs*) para convertirlos en productos que a su vez venderán en el mercado (*outputs*). La cantidad que pueden producir las empresas a partir de un vector determinado de *inputs* depende de la tecnología de producción que tengan a su disposición. Se supone que el conjunto de combinaciones factible de *inputs* y *outputs* es convexo.⁴ Por cada vector de precios, cada empresa elige un vector de *inputs* (compras) y un vector de *outputs* factibles (ventas) con vistas a maximizar sus beneficios.

Un equilibrio es un vector de precios tal que cuando todas las compras y las ventas de la economía se suman, la oferta es igual a la demanda en cada uno de los mercados. Gran parte de la investigación relativa al equilibrio general realizada hasta mediados la década de 1990 se dedicó a demostrar la existencia de un vector de precios de equilibrio dentro de los supuestos más débiles posibles relativos a las preferencias individuales y a la tecnología de producción. En efecto, el conjunto de decisiones individuales tomadas por las personas y las empresas de manera egoísta, sin ningún tipo de coordinación entre ellas, puede resultar factible y no generar desorden. Además de demostrar que el concepto de equilibrio no era vacío (los equilibrios siempre existen), los teóricos del EG también encontraron condiciones bajo las que este concepto no era excesivamente laxo: los equilibrios son determinados, con lo que queda excluida la posibilidad de un continuo de equilibrios.

Los resultados más notables de la teoría de EG, los *dos Teoremas Fundamentales de la Economía del Bienestar*, demuestran que estos equilibrios del mercado tienen interesantes propiedades de eficiencia. W. Pareto define un requisito básico de eficiencia que se ha vuelto esencial en la economía: una situación es eficiente, según Pareto, si no hay ninguna reasignación de los recursos económicos que permita mejorar el bienestar de un individuo sin perjudicar a otro. En esta noción no está presente en absoluto la justicia distributiva. Un reparto en el que una persona es dueña de todo mientras que el resto se muere de hambre es eficiente, siempre y cuando las preferencias individuales no se sacien.

El Primer Teorema Fundamental establece que todos los equilibrios competitivos son eficientes en el sentido de Pareto. Por consiguiente, el intercambio que tiene lugar en los mercados entre individuos que sólo tienen en cuenta sus intereses también hace un uso eficiente de los recursos existentes. El Segundo Teorema Fundamental dice que toda asignación eficiente de bienes se puede poner en práctica en forma de equilibrio competitivo, con una redistribución adecuada de los recursos iniciales. Continúa diciendo que este resultado implica que una economía socialista planificada no puede funcionar mejor que

1 Esta cuestión (la de los mercados frente a la planificación socialista) fue central en el debate político durante la Guerra Fría.

2 Este supuesto resulta más verosímil cuando todos los participantes son tan pequeños que con sus decisiones no pueden ejercer ninguna influencia sobre los precios. Es evidente que éste no es el caso en el mundo real. Los gobiernos han creado agencias reguladoras de la competencia que funcionan con mayor o menor eficacia y que intentan evitar que las empresas actúen en colusión manipulando los precios. Los trabajadores, por su parte, están en su mayor parte afiliados a sindicatos que les garantizan subidas salariales y mejora de las condiciones laborales.

3 Si se prefiere estrictamente la cesta de bienes A a la B, entonces cualquier combinación lineal convexa $\lambda B + (1 - \lambda) A$, ($\lambda > 0$) se prefiere estrictamente a B.

4 Si las combinaciones A y B son factibles, también lo son $\lambda B + (1 - \lambda) A$, $0 \leq \lambda \leq 1$.

los mercados competitivos, tras una redistribución apropiada de los recursos de carácter único.

Cuánto hay que redistribuir y cómo hacerlo sin perturbar el funcionamiento de los mercados es claramente una cuestión complementaria a la teoría de EG. Éste es el tipo de cuestiones que estudia la *Economía del Bienestar*. El gobierno sólo puede elegir si existen unas *preferencias sociales* que ordenan las políticas alternativas de acuerdo a la deseabilidad social de sus consecuencias. Ya en 1951, K. Arrow (premio Nobel de Economía en 1972) demostró que no era posible agregar las preferencias individuales en un orden de preferencias sociales que cumpliera una serie de requisitos razonables. La economía del bienestar, al igual que la economía pública, acabó por suponer que de un modo u otro las prioridades sociales podían quedar recogidas en una *función de bienestar social*. La actuación del gobierno se concibió entonces de forma análoga a las decisiones individuales: maximizar el bienestar social dentro de lo que resulta factible. Las aportaciones de P. Diamond y J. Mirrlees (premios Nobel de Economía, 1996) a mediados de la década de 1970 sentaron las bases de la economía pública moderna, arraigando con firmeza la teoría de la intervención gubernamental sobre los cimientos de la teoría del EG.

Este resumen, por supuesto, sólo hace referencia a los logros más importantes de la teoría del EG y la economía del bienestar.⁵ En la reunión de enero de 1994 de la Econometric Society, uno de los investigadores que más se han distinguido en el desarrollo de la teoría del EG, Andreu Mas-Colell, dictó una conferencia titulada «El futuro del equilibrio general».⁶ De esta presentación se desprendía la percepción de que la teoría del EG había alcanzado su punto álgido y que los investigadores jóvenes ya estaban estudiando cuestiones que hasta entonces se habían dejado de lado. Vamos a examinar algunas de estas nuevas líneas de investigación. Pero, antes de continuar con el tema central de mi artículo, es importante destacar una aportación fundamental a la teoría del EG: el rigor matemático. Ésta es, precisamente, la frase con la que Mas-Colell (1999) concluye su exposición: «Espero que las demostraciones mediante teoremas sigan teniendo un papel importante, si no tanto como hasta ahora —pues es posible que nos hayamos excedido— al menos sí una presencia considerable».

Principales puntos en los que la investigación se ha apartado recientemente del modelo estándar

Todos los teóricos del modelo de EG sabían que sus supuestos eran extremadamente rígidos, pero se consideraba que éste era el precio que había que pagar por disponer de un modelo nítido del funcionamiento de los mercados.

Una de las objeciones más obvias es que muchos mercados no tienen un número suficiente de empresas que justifiquen el supuesto de comportamiento competitivo. Hay situaciones en las que existe un monopolio, y hasta hay

otras en las que se considera *natural* el monopolio, como son el caso del suministro eléctrico, la televisión por cable, etcétera. Un monopolio no es *un agente económico sin influencia en el precio*, y puede que tenga en cuenta que el precio de venta dependerá de su volumen de producción. En este caso el equilibrio del mercado no sería eficiente en el sentido de Pareto. Lo mismo ocurre si hay varias empresas en el mercado, pero no las suficientes como para que cada una de ellas actúe como si su comportamiento no tuviera efecto en los precios. Esto ha dado origen al campo de la *organización industrial*, que se promovió sobre todo a finales de la década de 1980 y en la de 1990.⁷

Otra cuestión importante en la que la investigación se ha apartado del modelo tradicional de EG es la referida a la función de la información. Ello ocurrió en los ochenta y noventa del siglo pasado. En el modelo estándar se presupone que todos los participantes tienen acceso a la misma información, lo que podría querer decir que conocen las probabilidades relevantes en caso de incertidumbre. Sin embargo, es evidente que esto no siempre es así. El ejemplo tradicional (Akerlof, premio Nobel, 2001) es el del mercado de coches de segunda mano: el vendedor, que ha sido dueño del coche, tiene más información sobre su calidad que el comprador. En esta situación se puede demostrar que los equilibrios no son eficientes en el sentido de Pareto. El ejemplo de los coches usados de Akerlof (su famoso modelo de «el mercado de limones») es una parábola aplicable a una multitud de situaciones en las que la información no es simétrica. Hay otros ejemplos, como el mercado de seguros —en el que el asegurado puede tener más información sobre el riesgo que el asegurador—, los contratos laborales —en los que el *agente* empleado puede tener más información que su *empleador*—, y otros más (Akerlof, Mirrlees, Stiglitz, Vickrey, todos premios Nobel de Economía).⁸

Otro de los aspectos importantes en que la investigación se ha apartado del modelo tradicional tiene que ver con *externalidades*, concretamente me refiero a las situaciones en las que el consumo por parte de un agente pueda tener un efecto directo en el bienestar de otro. En especial, los casos que conciernen a bienes públicos, esto es, bienes que pueden utilizar muchos individuos de manera simultánea, como los hospitales, el transporte, la educación, la defensa, entran dentro de esta categoría. De nuevo, se puso de manifiesto que no se podía depender de los mercados competitivos para obtener una asignación de riqueza eficiente en el sentido de Pareto en estas situaciones.

Por ello se pasó a examinar todas estas cuestiones utilizando la teoría de juegos, que empezó usándose a principios del siglo xx para analizar la estrategia en juegos de salón. En 1940, O. Morgenstern y J. von Neumann escribieron la primera obra sobre teoría de juegos, que, afirmaban, era indicada para estudiar todas las situaciones económicas y sociales. Este enfoque pronto lo perfeccionó J. Nash, según el cual todas las situaciones deberían analizarse a partir del punto de vista de la decisión individual. Nash

⁵ Véase Mas-Colell et al. (1995), para una presentación del estado actual de la teoría del EG y de la microeconomía.

⁶ Véase Mas-Colell (1999).

⁷ Las obras de Tirole (1988) y Vives (2001) presentan una visión general muy completa sobre el tema.

⁸ Véase la presentación panorámica de Stiglitz (2002) en su discurso de aceptación del Premio Nobel.

9

Otros autores de importantes aportaciones a la teoría de juegos son L. Shapley, J. Aumann, y R. Selten (estos dos últimos también son premios Nobel de Economía junto con J. Nash y J. Harsanyi).

10

Nótese que estoy pasando por alto los puntos de vista nuevos relativos al comportamiento de las empresas más allá de la maximización de beneficios. Como excusa, citaré a Mas-Colell (1999) que señaló: «No estoy seguro de que [...] en última instancia, acabemos por considerar la teoría de la empresa como parte del núcleo de la ciencia económica».

11

El primer modelo clásico de democracia se lo debemos a Downs (1957).

12

Roemer (2001) ha escrito una obra rigurosa y exhaustiva sobre competición política.

13

Véase la obra de Persson y Tabellini, de gran influencia en este campo.

14

Véase Grossman y Helpman (2001).

15

Para un modelo general de conflicto véase Esteban y Ray (1999). Fearon (1996), Powell (1999) y Ray (2008b) han desarrollado argumentos para explicar por qué la sociedad puede fracasar a la hora de acordar un nuevo contrato social que para todos sea preferible a un costoso conflicto civil, que Pareto denomina «el elevado coste de los conflictos civiles».

16

Como consecuencia, el modelo de EG no es capaz de analizar cómo la economía puede alcanzar el vector de precios preciso que equilibraría los mercados.

17

Ray (1998) es el texto avanzado básico sobre economía del desarrollo. Véase también Ray (2008a) donde aparece una visión de conjunto de las novedades en este tema.

propuso la noción de *equilibrio*, que en la actualidad lleva su nombre (el equilibrio de Nash), en la que es necesario que cada decidor individual elija la mejor forma de proceder tomando en consideración las decisiones de los demás.

La teoría de juegos resultaba perfecta para analizar los mercados no competitivos, pero sólo tras las aportaciones de J. Harsanyi se puso de manifiesto que también era apropiada en las situaciones en las que la información era asimétrica, lo que hizo que se convirtiese en el método natural de examen de externalidades y bienes públicos. De esta forma pasó a ser la herramienta habitual en los análisis de la teoría microeconómica. A partir de mediados de la década de 1970, la teoría de juegos pasó a dominar la teoría económica. En las últimas décadas también ha demostrado ser una herramienta importante en la macroeconomía e incluso en ciencia política. Daba la impresión que cualquier problema que surgiese en las ciencias sociales podía considerarse como una aplicación de la teoría de juegos.⁹

La investigación actual, sin embargo, indica que, o bien no todos los problemas importantes pueden resolverse aplicando la teoría de juegos, o bien ésta no proporciona un análisis completo de cada situación. Los diferentes puntos en los que la investigación reciente se ha apartado de este paradigma pueden clasificarse en tres categorías: (i) la manera en que los individuos y las empresas toman decisiones; (ii) la forma en que lo hacen los gobiernos; y (iii) la manera en que interactúan los agentes.

A pesar de tratar un área tan específica de la ciencia económica, existen demasiadas líneas de investigación como para realizar una presentación coherente y exhaustiva. Por lo tanto, me centraré únicamente en la investigación de frontera relativa a la elección racional del individuo.¹⁰ No obstante, antes de continuar, daré una visión esquemática de los principales avances referidos a la toma colectiva de decisiones y a la desviación de la hipótesis de comportamiento competitivo.

Está claro que las políticas económicas de los gobiernos no se deciden con el propósito de maximizar unas preferencias sociales. Más bien, en las democracias, los partidos políticos proponen medidas y los ciudadanos votan de acuerdo con los programas electorales presentados.¹¹ Ésta es la forma en la que las democracias han decidido resolver los intereses opuestos que caracterizan a todas las sociedades. Por consiguiente, si queremos comprender los programas políticos en vigor, tenemos que explicar cómo los partidos eligen sus programas y cómo se comportan una vez que están en el gobierno.¹² Este enfoque se conoce con el nombre de *economía política positiva*.¹³ Por supuesto, existe también abundante literatura sobre la influencia que ejercen las grandes corporaciones y los grupos de presión en la toma de decisiones de los gobiernos.¹⁴

El sistema político es en sí mismo endógeno y a menudo responde a la configuración de los intereses confrontados en el seno de la sociedad. Acemoglu y Robinson (2006)

han estudiado cómo los cambios sociales y económicos han forzado la adaptación del marco político. En concreto, señalan que las democracias han evolucionado como un mecanismo de compromiso: la tensión por la desigualdad en la distribución de la riqueza y la amenaza de revolución obligó en su tiempo a los monarcas a compartir sus recursos. Sin embargo, dado que la promesa de repartir recursos puede no resultar creíble por sí misma, se hizo necesario garantizar el voto democrático a amplios sectores de la sociedad para hacer verosímiles estos compromisos, evitando así posibles conflictos. Está claro que el contrato social no siempre puede modificarse en respuesta a los cambios sociales. En la segunda mitad del siglo xx más de 16 millones de ciudadanos murieron en guerras civiles, una cifra cinco veces superior a la de muertos en el campo de batalla durante el mismo periodo. La amenaza de una guerra civil y la inestabilidad política derivada de ella en los países en desarrollo es hoy por hoy tan endémica que el Banco Mundial la considera el impedimento más grave para conseguir que la ayuda extranjera sea eficaz a la hora de fomentar el crecimiento en dichos países. En el terreno de la ciencia económica, Hirshleifer (1991), Grossman (1991) y Skaperdas (1992) fueron los pioneros en el análisis del conflicto, pero aún estamos muy lejos de entender de manera satisfactoria las causas que provocan una guerra civil.¹⁵

El modelo de EG no explica en absoluto la manera en la que los agentes interactúan. Cómo los productos manufacturados por las empresas llegan hasta los consumidores o a otras empresas queda fuera del modelo.¹⁶ Esta falta de interés en el funcionamiento efectivo de las transacciones económicas se debe en parte a que el modelo se centra exclusivamente en los mercados competitivos en equilibrio. En efecto, si los mercados son competitivos y están en equilibrio no tiene sentido explorar si existen transacciones mejores. La demanda (quizás, por arte de magia) se satisface, y no es posible encontrar un vendedor que ofrezca un precio más bajo. Sin embargo, si los mismos bienes tienen precios diferentes o si algunas necesidades (como la de un individuo que quiera vender su trabajo) no quedan satisfechas, entonces se vuelve importante saber cómo funcionan de hecho las transacciones.

El mayor esfuerzo por investigar el funcionamiento real de las interacciones económicas ha sido realizado por la *Economía del Desarrollo*. Sólo con observar de manera superficial los países subdesarrollados resulta evidente que sus economías no se pueden concebir como un conjunto de mercados competitivos en equilibrio. Debemos a esta área de la economía la mayoría de las reflexiones sobre el papel de las instituciones, de las conexiones personales y de las normas sociales en encauzar las interacciones económicas.¹⁷ En un plano más formal, la existencia de conexiones sociales y su efecto en las interacciones económicas ha llamado la atención de los teóricos económicos. La esencia de este modelo tiene elementos comunes con los modelos de *interacción local* en el campo de la física, aun-

que con una particularidad: los nodos aquí son individuos o empresas que buscan el mayor beneficio posible y que tienen capacidad de crear o romper lazos.¹⁸ Este enfoque ha permitido un mayor entendimiento en diferentes áreas económicas como la educación¹⁹ y el mercado laboral.²⁰

La teoría clásica de la elección racional

Muchas de las nuevas tendencias dentro de la investigación en economía están motivadas por la necesidad de enraizar de manera más sólida la teoría con los hechos. Esto es especialmente cierto en lo que se refiere a la *economía conductista*, como algunos colegas la han denominado. Desde que se observó que el comportamiento de los individuos está frecuentemente en desacuerdo con las decisiones que se desprenden de los supuestos del modelo de la elección racional, los economistas han empezado a fijarse en los patrones de comportamiento que los psicólogos han identificado por medio de experimentos controlados de laboratorio. La labor vanguardista de los psicólogos Kahneman (premio Nobel de Economía 2002) y Tversky²¹ ha tenido una enorme influencia en la ciencia económica reciente. Además de ampliar la perspectiva de los economistas, las conclusiones a las que han llegado los psicólogos han favorecido el auge de experimentos sobre el comportamiento individual y colectivo. C. Camerer, E. Fehr, D. Laibson y M. Rabin, por mencionar unos pocos, son algunos de los economistas que más intensamente han trabajado en este terreno.²²

Con el fin de poner un poco de orden, considero que es útil separar los ingredientes fundamentales en la toma de decisiones por parte del individuo.

El primero es el *input* informacional. Los individuos observan cierta información sobre el estado del mundo. En la teoría general del consumidor, esta información se refiere a los precios y a la renta y, en un entorno de incertidumbre, a la probabilidad de cada uno de los posibles Estados del mundo. Este tipo de información normalmente limita las acciones que resultan accesibles a cada individuo.

El segundo ingrediente es el cálculo de las consecuencias (deterministas o probabilistas) que traerá consigo cada una de las acciones posibles. Por ejemplo, decidimos trabajar ocho horas y comprar carne y pescado. O podemos decidir emplear 10 euros en comprar un billete de lotería con unas probabilidades determinadas de ganar un premio, en lugar de destinarlos a consumir algo.

El tercer ingrediente consiste en las preferencias individuales. Se supone que éstas dan lugar a una ordenación de todas las posibles consecuencias (independientemente de las acciones en sí mismas) de acuerdo con lo deseables que sean, como se explicó anteriormente. Cuando las consecuencias son probabilistas, los individuos clasifican las acciones según su *utilidad esperada*, es decir, la media ponderada de la utilidad de los distintos resultados, utilizando las probabilidades como ponderaciones. En el

modelo estándar se presupone que estas preferencias son egoístas, y que no tienen en cuenta lo que otros puedan obtener.

El último ingrediente es la elección racional. Con ello se quiere decir que cada individuo es capaz de resolver el problema de la maximización restringida, que consiste en determinar la acción factible que tenga la consecuencia más deseable.

La psicología y las decisiones del individuo

La labor experimental sobre el comportamiento individual consiste en poner a un grupo de individuos ante una situación (lo más controlada posible) en la que la teoría tradicional de la decisión haya formulado una predicción inequívoca, de manera que el investigador pueda contrastar las elecciones reales de cada uno de ellos con las que se pronosticaron. Hay una gran variedad de experimentos de este tipo que estudian diferentes casos en los que hay una transgresión de las predicciones que aparecen en el modelo general.²³

A título de ejemplo, uno de los experimentos más comunes consiste en someter a un grupo de individuos al llamado «juego del ultimátum». Se empareja a los individuos al azar, y a uno se le da el papel del que propone y al otro de receptor de la propuesta. El primero propone una división de una cantidad determinada de dinero entre los dos jugadores. A continuación, el segundo acepta la propuesta, con lo que el dinero quedará dividido de acuerdo con la asignación sugerida, o la rechaza, con lo que ninguno de los dos jugadores recibirá nada. Si los jugadores actúan únicamente en beneficio propio, el segundo jugador aceptará cualquier cantidad de dinero que sea positiva. Sabiendo esto, un proponente egoísta dará una cantidad arbitrariamente pequeña al receptor y se quedará con el resto. Sin embargo, resulta que en todas las modalidades de este experimento hay una proporción considerable de proponentes que sugieren dividir el dinero de manera bastante equitativa, así como de receptores que están incluso dispuestos a renunciar a cantidades nada desdeñables con tal de *castigar* las propuestas injustas.

El experimento anterior no es más que un ejemplo más de la enorme cantidad de patrones de elección que en la actualidad analizan los economistas. Algunos de estos experimentos ponen a prueba determinados elementos específicos del proceso, ya descrito, de toma de decisiones. Pero otros no están tan nítidamente dirigidos y simplemente intentan identificar casos en los que no se cumplen las predicciones de la teoría estándar de elección racional.

Estudiemos algunos de los aspectos complejos más relevantes que dicha teoría elimina por hipótesis en cada uno de los ingredientes de la toma de decisión. Algunos de estos aspectos han sido ya examinados por la economía conductista, pero quedan aún muchos para estudiar rigurosamente. Como quedará claro, mi postura es, en cierto modo, ambivalente. Por un lado, considero que la ciencia económica debe ampliar su modelo básico de elec-

18
Véase Jackson (2008b), que incorpora un estudio sobre las aportaciones recientes, así como los libros de Goyal (2007) y Jackson (2008a) que incluyen una amplia presentación del tema.

19
Véase Benabou (1993, 1996).

20
Véase el estudio de Calvo-Armengol y Yoannides (2008).

21
Tversky murió en 1996. Véase el ensayo de Laibson y Zaeckhouser sobre sus aportaciones.

22
Las conferencias de Camerer y Rabin en el IX Congreso Mundial de la Econometric Society proporcionan una visión de conjunto del potencial de la combinación de psicología y economía (véase también los planteamientos de Ariel Rubinstein). Todos publicados en Bundell et al. (2006).

23
Véase Della Vigna (2008) para un panorama exhaustivo de los diferentes tipos de experimentos.

ción racional, pero por otro soy bastante escéptico, por no decir crítico, con respecto a muchas de las afirmaciones formuladas por la economía conductista. En este sentido me siento más en consonancia con las posturas críticas de Gul y Pesendorfer (2008) y Rubinstein (2006).

El primer ingrediente de la elección es el procesamiento de información. Hay muchos canales a través de los que la adquisición de información puede repercutir en las decisiones. En primer lugar, los individuos categorizan la información, tema que ha sido objeto de estudio por parte de psicólogos sociales durante las últimas cinco décadas, pero al que los economistas han empezado a prestar atención sólo recientemente. Es evidente que este proceso puede influir en nuestras decisiones. Fryer y Jackson (2008) estudian cómo un procesamiento eficaz de la información conduce a una categorización menos depurada de los tipos de experiencias que se observan con menor frecuencia, agrupándolas. Como consecuencia, las personas que toman las decisiones formulan predicciones menos precisas cuando abordan estos temas, lo cual puede dar lugar a una discriminación. En segundo lugar, los individuos deben tener una idea de qué información es relevante con respecto a la decisión en cuestión y cuál no lo es. En otras palabras, tienen que considerar un *modelo* que relacione las posibles acciones con sus consecuencias, como explicaremos más adelante. No obstante, hay evidencia psicológica de que los individuos tienden a censurar la información que refuta su visión del mundo. Es más, Bebabou y Tirole (2006) mantienen que este mecanismo de censura puede apoyar la creencia de que el grado de movilidad social es mayor del que en realidad es, y de este modo los individuos se auto-inducen a realizar un esfuerzo mayor que el que justificaría la elección racional. Estas creencias tan poco realistas son sobre todo necesarias en países en los que la red de prestaciones sociales es limitada, como Estados Unidos. Por último, es posible que el supuesto de elección racional relativo a que los individuos utilizarán la información para realizar actualizaciones bayesianas de las probabilidades relevantes esté injustificado. Como señalan Gilboa et al. (2007), se dan casos en los que, para empezar, los individuos no pueden tener creencias previas representables por probabilidades, porque las creencias racionales y justificadas no consiguen precisar una probabilidad numérica.

El segundo ingrediente consiste en identificar las consecuencias que se derivan de las acciones, ya sea de manera determinista o probabilista. Este paso supone que, a la vista de la evidencia, los individuos pueden identificar un modelo que encaja con la información, el cual es único. Es indudable que, en líneas generales, esto no es así. De una manera más formal, Aragonés et al. (2005) señalan que con una base de conocimiento determinada, encontrar un pequeño grupo de variables con el que conseguir un valor concreto de R^2 es *computacionalmente difícil*, en el sentido informático del término. Por ello las afirmaciones retó-

ricas que no aportan nueva evidencia pueden dar lugar a cambios en las decisiones, ya que pueden hacer que uno sea consciente de determinadas relaciones entre variables conocidas, que resultan obvias a posteriori, pero en las que no se pensó antes. Múltiples teorías y múltiples decisiones son compatibles con una evidencia dada. Picketty (1995) elaboró un modelo en el que la creencia en una teoría específica induce a tomar decisiones que generan pruebas que la confirman. Individuos diferentes pueden mantener teorías distintas, y los padres tienen el incentivo de transmitir sus teorías a sus hijos. Por último, otra forma de toma de decisiones en los casos en los que no hay una interpretación determinada de la evidencia es a través de la imitación social. Banerjee (1992) elaboró un modelo de *gregarismo* en el que los individuos sacan conclusiones a partir de la observación del comportamiento de otros en lo que respecta a la información que puedan tener, lo cual les hace actuar de forma similar.²⁴ Una corriente alternativa es el efecto que tiene la identidad social en el comportamiento. Akerlof y Kranton (2000) fueron los primeros que despertaron la atención de los economistas sobre la identificación individual con categorías o prototipos determinados. En mi opinión, se trata de una línea de investigación importante, aunque desgraciadamente poco explorada todavía.²⁵

El tercer ingrediente es el de las preferencias individuales. Quizás éste es el aspecto en el que la teoría económica tradicional se muestra más limitada y menos realista. En primer lugar, supone que las preferencias vienen definidas únicamente por el consumo personal, lo que excluye las motivaciones altruistas, de las cuales existen abundantes pruebas. Son numerosísimas las publicaciones sobre altruismo, sobre todo el basado en la reciprocidad. Rabin (1993), Levine (1998), Fehr y Schmidt (1999), Bolton y Ockenfels (2000) y Falk y Fischbacher (2006) se han centrado en la interacción entre parejas de jugadores en las que cada jugador intenta deducir las motivaciones de su pareja y en función de lo observado modifica el peso que da a los demás en sus preferencias, concediendo mayor peso a los que considera altruistas o benévolos y menor a los que son egoístas o malévolos. Además, puede que los individuos valoren las acciones por sí mismas, con independencia de sus consecuencias. Ello tiene que ver con uno mismo (autoestima/amor propio) y con los demás (juicios éticos). Para un análisis económico en el que los padres intentan inculcar hábitos de trabajo a sus hijos motivados por un sentimiento de culpa véase Lindbeck et al. (2006), y Tabellini (2007), para temas relativos a la adopción y transmisión de valores morales universales.²⁶ Por último, los individuos parecen experimentar cambios en sus preferencias tanto con el paso del tiempo²⁷ como a través de limitaciones en su abanico de opciones.²⁸

El cuarto ingrediente es la toma de decisión propiamente dicha, que consiste en el proceso de combinar toda la información disponible y convertirla en la elección

24

Véase la reciente panorámica sobre la conducta gregaria de Rook (2006).

25

Esteban y Ray (1994) presentaron la idea de que el conflicto social se ve impulsado por el sentido de identidad con el propio grupo, combinado con sentido complementario de alienación con respecto a los otros. Simplemente axiomatizaron un índice de *polarización* pero no desarrollaron un modelo de formación de identidad.

26

En lo que se refiere a las normas relativas al trabajo, las aportaciones fundamentales de Moffit (1983) y Besley y Coate (1992) plantean el caso en el que hay un estigma asociado a vivir de prestaciones sociales. Lindbeck et al. (1999) han ampliado este análisis y han incluido el voto sobre las prestaciones sociales. Cervellati et al. (2008) deja que los juicios morales sobre el esfuerzo determinen la autoestima y la estima hacia los demás.

27

El coste psicológico del aplazamiento de una recompensa determinada es más elevado si se produce inmediatamente que si ocurre más tarde. Laibson (1997) ha acuñado el término *descuento hiperbólico* para designar esta variación en la tasa de impaciencia.

28

He aquí un ejemplo típico: sin duda prefiero no fumar. También me gustaría comer con una amiga que da la casualidad de que fuma. Puedo comprometerme a no fumar evitando comer con mi amiga. Sé que si como con ella tendré tentaciones de fumar. Gul y Pesendorfer (2001) han estudiado esta idea en profundidad.

de una única acción. El supuesto de racionalidad da por hecho que los individuos elegirán la acción que consideren mejor dentro de todas las posibilidades a su disposición. Ello quiere decir que el individuo no se comportará racionalmente sólo cuando el analista le ofrezca una opción diferente de la que ha elegido y el individuo reconozca que en efecto es preferible.

La consideración conjunta de la información disponible, el vínculo entre las acciones y las consecuencias y la evaluación de estas últimas exigen una capacidad de razonamiento considerable, la cual depende de la educación, la formación y las experiencias pasadas. Por ello podemos concluir que los individuos se comportan de manera no racional sólo cuando se empeñan en mantener sus elecciones después de que un analista les haya mostrado que pueden elegir algo mejor. Volveremos a tratar esta noción de racionalidad basada en Gilboa y Schmeidler (2001).²⁹ En caso de certidumbre, la elección racional viene a ser lo mismo que resolver un problema de maximización bajo restricciones de factibilidad. Sin embargo, ya en la década de 1950, Simon (premio Nobel 1978) mantuvo que la racionalidad individual estaba *acotada*, en el sentido de que la capacidad de cálculo humano es limitada. Ahora bien, nótese que la *racionalidad acotada* en tanto limitación de cálculo no implica ausencia de racionalidad en el sentido anteriormente explicado.

El proceso de razonamiento necesario para tomar una decisión en caso de incertidumbre es mucho más complejo, ya que exige además tener en cuenta las probabilidades de cada una de las posibles consecuencias. La teoría tradicional de la decisión racional presupone que los individuos evalúan cada actuación mediante la suma ponderada de las valoraciones de sus consecuencias, utilizando las probabilidades relevantes como ponderaciones.³⁰ La evidencia empírica parece confirmar que algunos de los axiomas se incumplen de manera regular, las llamadas paradojas de Allais y Ellsberg. Con objeto de reconciliar la teoría y el comportamiento, Kahneman y Tversky (1979) propusieron la *teoría de las perspectivas* en la toma de decisiones bajo situaciones de riesgo. Basándose en el comportamiento, y no en axiomas, defendieron que las probabilidades no entran en el proceso de evaluación de una acción de manera lineal, sino a través de una función de ponderación que exagera las probabilidades más bajas y modera las más elevadas. Asimismo, la evaluación de cada consecuencia se mide como una desviación respecto a un resultado de referencia.³¹ Nótese que este enfoque sigue presuponiendo que existen probabilidades bien definidas para cada posible consecuencia de cada acción. Ahora bien, esto es muy poco frecuente. Gilboa y Schmeidler (2001) sostienen que las decisiones se basan en los resultados que se han observado en casos anteriores y que se consideran *similares* al problema actual. A partir de un conjunto de axiomas, deducen que la evaluación de una acción es la suma de los niveles de utilidad que resultaron

de usarla en casos anteriores, ponderados por el grado de similitud con el problema actual.

Por lo que se refiere a la elección racional, podemos concluir que, mientras que la toma de decisiones racionales en casos de certidumbre no resulta muy polémica, en los casos de incertidumbre todavía es un tema abierto. En suma, hemos visto que el comportamiento real parece mostrar desviaciones en cada uno de los elementos que forman parte del proceso de toma de decisiones. Sin duda, todavía se puede enriquecer nuestro modelo con respecto a cómo se adquiere y se procesa la información, a cómo los individuos relacionan las acciones con sus consecuencias, y hasta a cuáles son las diferentes dimensiones que los agentes valoran (además de los bienes materiales objeto de comercio). No obstante, ninguno de estos cambios parece tener mucha relación con la noción básica de tomar una decisión óptima bajo determinadas restricciones.

En el próximo apartado analizaremos qué es lo que podemos aprender de los resultados que hasta ahora ha obtenido la economía conductista y hasta qué punto suponen una dificultad para el supuesto de la toma racional de decisiones por parte del individuo.

Economía conductista. Balance general

Cabe preguntarse adónde nos lleva el examen de los vínculos existentes entre psicología y ciencia económica. Algunos economistas conductistas han considerado que la amplia y sólida evidencia empírica de que una serie de factores psicológicos son importantes en la toma de decisiones por parte de los individuos pone a prueba el paradigma central de la ciencia económica relativo a la elección racional.³²

Sostendré que no está claro cómo puede la evidencia empírica extrapolarse fuera del laboratorio (y qué se puede aprender de ello), y que es probable que el enriquecimiento de la descripción del comportamiento de los decisores tenga mayor influencia en los modelos de economía aplicada que en el paradigma y supuesto central según el cual los individuos actúan de manera racional y, esencialmente, movidos por su propio interés.

¿Qué podemos aprender de la evidencia empírica?

En primer lugar, confirmamos de forma controlada que los individuos se comportan de manera diferente a como prescribe la teoría económica tradicional. En efecto, a los individuos les importan otras cosas aparte del consumo material, como por ejemplo, las acciones que pueden adoptar en sí mismas, sus propios juicios morales y los de los demás, etcétera. Quienes toman decisiones tampoco procesan la información perfectamente, y a menudo no cumplen alguno de los axiomas del comportamiento racional. No obstante, no siempre es obvio cómo se debe interpretar esta evidencia, ni, aunque fueran inequívocas, si estas transgresiones que se han encontrado en experimentos deberían formar parte del modelo teórico general de comportamiento del individuo.

29

Para un análisis más detallado véase Gilboa et al. (2008).

30

Para ser más precisos, el supuesto es que las elecciones del individuo respetan los axiomas propuestos por Von Neumann y Morgenstern (1944), lo que significa que la evaluación de una actuación es su utilidad esperada.

31

El caso de la paradoja de Ellsberg se examina en Fox, C. R. y Tversky (1995).

32

Rabin y Thaler (2001) califican a la teoría tradicional de la utilidad esperada de «loro muerto», inspirándose en un *sketch* de Monthly Python.

En cierto modo, el experimento de Kahneman y Tversky (1984) en el que queda demostrado que el *framing* (el encuadre de una situación) afecta al comportamiento del individuo, inspira cierto escepticismo con respecto a lo que se puede aprender de los experimentos. En primer lugar, existe la sospecha de que el comportamiento objeto de observación haya podido ser provocado por la manera concreta en la que el problema en cuestión se ha planteado a los participantes.³³ ¿Debemos, entonces, concluir que la gente por lo general no cumple los supuestos más básicos de esta teoría, o que, a veces, ante determinados *planteamientos ingeniosos*, puede actuar de forma muy irracional?

En la labor del economista conductista está implícita la creencia de que existe un *patrón natural* de comportamiento, que no quedó recogido debidamente en la teoría clásica de la toma de decisiones, y que se puede determinar mediante experimentos críticos.

Camerer y Loewenstein (2003) señalan que «la economía del comportamiento aumenta la capacidad explicativa de la ciencia económica, al proporcionarle unas bases psicológicas más realistas» y que «la convicción de que incrementar el realismo de los cimientos psicológicos del análisis económico hará que la ciencia económica mejore en sí misma». Asimismo, Rabin (1998) afirma que «puesto que la psicología examina de manera sistemática el criterio, el comportamiento y el bienestar humanos, nos puede enseñar que los seres humanos son diferentes a como la ciencia económica los ha descrito tradicionalmente». Por tanto, el propósito es comprender la naturaleza *auténtica* de la toma de decisiones de los individuos a partir de observaciones factuales en experimentos o por otros medios.³⁴

Cabe preguntarnos si es posible captar dicha *naturaleza* de la toma de decisiones mediante experimentos o, incluso, si ésta existe de una forma significativa.

Al margen de las reservas que existan sobre la posibilidad de controlar eficazmente los experimentos, aún no está clara cuál es la *naturaleza* exacta que estamos midiendo. Con objeto de ilustrar mi argumento, tomemos el experimento más habitual y que ya hemos descrito aquí: el juego del ultimátum. El rechazo a las propuestas *injustas* se interpreta como una muestra de que a los individuos también les importan otras cosas que no sean las compensaciones monetarias personales.³⁵ Sin embargo, también puede tratarse de una reacción emocional que, momentáneamente, impide interpretar claramente los dictados de la razón. El grado hasta el cual la razón domeña las emociones varía según los diferentes individuos, posiblemente dependiendo de su educación y formación, y, en cualquier caso, sólo los dictados de la razón deberían considerarse parte del comportamiento racional. Si lo que nos interesa son las elecciones que un grupo determinado de individuos pueda tomar en unas circunstancias concretas, puede que sea fundamental saber si van a reaccionar emocionalmente o si lo van a pensar fríamente.³⁶ Sin embargo, parece lógico que una teoría general sobre el comportamiento del indivi-

duo deba abstraerse del hecho de que, de manera momentánea, podamos desviarnos del comportamiento racional.

Esto plantea una cuestión fundamental sobre la que deberemos volver: si la racionalidad es algo positivo o normativo. ¿Debe la sociedad enseñar a los ciudadanos a ser racionales?³⁷ De hecho, es lo que intenta hacer, hoy por hoy, el sistema de educación obligatoria.

Hasta para un nivel dado de capacidad de razonamiento complejo, las experiencias concretas o la formación pueden tener un efecto profundo en el comportamiento. El trabajo de Amiel y Cowell (1992), en el que se intenta de manera empírica determinar las nociones de equidad que, de hecho, utilizan los individuos, viene muy al caso para corroborar mi argumento. Se mostró a los estudiantes dos listas de (diez) rentas y tenían que decidir cuál de las dos era más desigual. El estudio tenía como objeto determinar cuál de los diferentes criterios de equidad empleados en economía obtenía una mayor aceptación. Entre estos criterios se analizó el *principio de transferencias progresivas* popularizado por Atkinson (1970), según el cual, si transferimos un euro de una persona a otra más pobre la distribución resultante es menos desigual. El resultado que tiene importancia aquí es que este principio tuvo una amplia aceptación entre los estudiantes de Económicas, que de manera directa o indirecta estaban familiarizados con este concepto, y muy baja entre otros estudiantes. En efecto, podemos interpretar información con más facilidad cuando se nos ha explicado cómo organizarla.

La pregunta anterior de si existe una *naturaleza* de la toma de decisiones que se pueda captar mediante experimentos era, en cierto modo, retórica, pues hoy por hoy nuestros conocimientos aún no permiten llevar a cabo experimentos críticos. Como han estudiado Levitt y List (2007) de forma detallada, hasta los experimentos diseñados con el mayor detenimiento no pueden garantizar al analista un control eficaz de todos los factores ajenos que pueden intervenir. Por consiguiente, si bien los experimentos son muy útiles a la hora de identificar las desviaciones del comportamiento predecible, no pueden, por lo general, determinar de forma inequívoca las causas por las que éstas se producen. En mi opinión es muy importante que se sigan realizando experimentos, pero considero que se trata de un proyecto a largo plazo que requiere tiempo, esfuerzo y paciencia.

Conductismo y elección racional

El conductismo será más influyente en modelos de economía aplicada que a la hora de redefinir el paradigma central de la elección racional del individuo. Presentaré dos argumentos que apoyan mi tesis.

El primero es que todavía hay demasiados aspectos de la conducta humana sobre los que tenemos una comprensión muy limitada. Sabemos, por ejemplo, que los individuos pueden estar motivados por sentimientos altruistas, pero todavía no podemos comprender lo que

33

Este y otros argumentos relativos a la dificultad de extraer conclusiones a partir de experimentos sobre el comportamiento se examinan detenidamente en Levitt y List (2007).

34

No voy a tratar los intentos actuales relacionados con los estudios sobre el vínculo entre la toma de decisiones y la actividad cerebral. Véase Gul y Pesendorfer (2008) para una visión crítica de conjunto.

35

Esta interpretación se ha visto reforzada por unos interesantes resultados, según los cuales, cuando el proponente es reemplazado por una máquina de la que se sabe que selecciona propuestas al azar, el segundo jugador se muestra más receptivo a propuestas injustas.

36

Hasta en los casos en que el rechazo se debe a una decepción moral, el que realiza el experimento debería comprobar si posponer la decisión de rechazar el dinero hasta el día siguiente alteraría los resultados.

37

«Los seres humanos, según mantenían los romanos, están hechos de dos elementos: un espíritu racional e inteligente y el cuerpo físico. [...] En las personas completamente racionales, como el caso de la élite romana, el espíritu racional controlaba al cuerpo físico. Pero en los seres humanos de menor categoría, los bárbaros, el cuerpo gobernaba la mente. [...] Así, mientras que los romanos calculaban las probabilidades, elaboraban planes sensatos y actuaban conforme a ellos en todo momento, los pobres bárbaros estaban siempre sujetos a los caprichos del destino». Heather P. (2005) *The Fall of the Roman Empire*, Oxford University Press (p. 69).

provoca un grado distinto de intensidad entre los miembros de la población. Algunos investigadores consideran que este altruismo está motivado por la búsqueda de los beneficios de la reciprocidad, los cuales pueden ser materiales o de actitud. Otros investigadores opinan que este altruismo tiene su origen en convicciones morales. También se ha observado que el grado de altruismo depende de la proporción del grupo de individuos que se comporta de forma altruista. Sólo podemos especular con respecto a cómo interactúan todos estos factores, pues hasta el momento ignoramos si los valores morales, las reacciones al comportamiento observado por parte de otros, la búsqueda de reciprocidad y otros aspectos similares son parámetros exógenos o tienen su origen al menos en parte en las variables que estamos intentando analizar. Es evidente que sin saber exactamente (o especular) qué condiciona qué, estas características del comportamiento no pueden incluirse en un modelo general.

Mi segundo argumento es que, aunque supiésemos mucho más sobre el comportamiento del individuo, debemos considerar cuántos atributos específicos queremos que recoja nuestro modelo. Cuando buscamos predecir cuál será la demanda de un producto determinado (un modelo nuevo de coche, por ejemplo) la teoría del consumidor que aparece en los manuales no resulta de gran ayuda. Los departamentos de ventas de las grandes empresas saben de sobra que existen muchas razones al margen del precio que hacen que la gente compre un producto, así como que un sector determinado de consumidores es receptivo al sentimiento de orgullo que produce conducir un coche nuevo, mientras que otro se dedica a leer detenidamente los informes de las asociaciones de consumidores, etcétera. Al interpretar correctamente la reacción de cada tipo de consumidor, pueden calcular la posible demanda con una precisión considerable. Ahora bien, la mayoría de los investigadores considera que este tipo de actividades no forma parte de la ciencia económica como tal.

Entonces, cabe preguntarnos cómo ha abordado la ciencia económica los factores que no encajan con los supuestos del modelo central. Durante mucho tiempo la ciencia económica moderna ha identificado *anomalías* tales como los bienes públicos (cuyo consumo no reduce las existencias, como la televisión pública o el orden público, como en Samuelson 1954), las inconsistencias en las elecciones intertemporales (analizadas por Strotz 1956) o la influencia del estatus social en el consumo (Duessenberry 1953). Sin embargo, el que se registrasen estas anomalías no menoscabó la teoría tradicional de la elección racional. Más bien la ciencia económica reaccionó elaborando modelos *auxiliares* que examinaran cómo estos casos que se alejaban de los supuestos tradicionales podían modificar las intuiciones derivadas del modelo de EG.

La teoría tradicional de la toma de decisiones no pretende ser descriptiva en el sentido literal de la palabra. La aportación del modelo de EG no ha sido formular teorías

que pronostiquen algo con precisión, sino más bien proporcionar una nueva visión del mundo que sea realmente esclarecedora. Renunciar a la precisión en aras de un mayor entendimiento es una de las concesiones habituales de la ciencia económica, y quizás de las ciencias sociales en general. Hasta qué punto podemos ignorar los atributos específicos relativos al comportamiento, es algo discutible. La pregunta sería si la investigación en economía debe continuar en la línea de lo explicado anteriormente, y elaborar también modelos *auxiliares*, al tiempo que se mantiene la esencia del modelo de EG como paradigma central. A día de hoy, mi respuesta es que sí, al menos mientras sigamos sin poder definir con precisión los factores exógenos que determinan los patrones de comportamiento observable.³⁸

¿Cuán realista debe ser una teoría?

No cabe duda de que, para poder llevar a cabo aplicaciones prácticas, interesa disponer de una teoría lo más precisa posible. Ahora bien, para las aplicaciones teóricas, tales como la formulación de teoremas del bienestar, no está claro que unos supuestos más precisos den como resultado una precisión también mayor, por no hablar de unas conclusiones más útiles. Ello se debe a que las aplicaciones teóricas utilizan modelos que se sabe son falsos como método para clarificar y verificar razonamientos. Determinados supuestos, que está claro que son incorrectos cuando se ponen a prueba en un laboratorio, pueden resultar más útiles para ciertos propósitos y menos para otros. Existe el peligro de que un resultado experimental como el del efecto de la manipulación de la información (*framing*), cuando se combina con otros supuestos teóricos, pueda conducir a un resultado menos realista que la hipótesis de que los efectos de la manipulación son irrelevantes.

Por consiguiente, lo que nos debemos preguntar al abordar la ciencia económica, no es si un supuesto determinado es preciso, sino si, como señaló Milton Friedman (premio Nobel 1976) hace ya bastante tiempo, nos lleva a conclusiones más precisas cuando lo combinamos con otros supuestos, y, lo que es más importante, si indica que hay un equilibrio razonable entre precisión y solidez. Si al final rechazamos todos los supuestos, y no llegamos a conclusión alguna, la precisión de nuestros modelos habrá servido de bien poco.

A modo de conclusión

Algunos investigadores han tenido la tentación de interpretar las desviaciones observadas en el comportamiento como un obstáculo al supuesto de racionalidad. Como hemos visto, muchas de ellas se deben bien a errores al procesar la información, al *framing* particular, a una mala comprensión de la relación entre las acciones y sus consecuencias, o bien a alteraciones temporales de las preferencias o a errores de cálculo (provocados por emociones y otros factores similares). Como señalaron Gilboa y Schmei-

38

Personalmente, considero que es más lamentable el supuesto de comportamiento competitivo cuando, por ejemplo, el mercado de petróleo lo controlan los países de la OPEP.

dlar (2001) y Gilboa et al. (2008), todas ellas tienen algo en común: enfrentados al análisis de su comportamiento, quienes han tomado la decisión cambiarían sus elecciones. Por ejemplo, eliminarían los errores identificables relativos a los razonamientos o a las reacciones precipitadas. Ello significa que lo que resulta irracional son los comportamientos que no resisten el análisis, los que es probable que se cambien al hablar con un experto o al intercambiar opiniones sobre la decisión con otros individuos en una situación parecida. Parece por tanto más útil centrarse en aquellas desviaciones de la teoría tradicional que pasan esta prueba de solidez, que se consideran *racionales* en este sentido. Otras transgresiones a veces nos hacen pensar y a menudo nos divierten, pero no pueden tenerse en cuenta como base de un análisis económico responsable.

Quisiera terminar este ensayo con una pequeña referencia al futuro inmediato de la investigación de la economía conductista. En el momento actual son muchas las cuestiones en las que la investigación se ha apartado del modelo clásico de elección racional, todas motivadas por la evidencia psicológica experimental. En palabras de Rubinstein (2008), un modelo «que tiene en cuenta factores como el sentido de la justicia, la envidia, la valoración del momento presente y otros similares, en la actualidad no sólo se permite sino que se prefiere». Toda esta variedad de casos en los que la investigación se ha apartado del modelo estándar resulta ciertamente atractiva desde un punto de vista intelectual, pero también produce desconcierto y sensación de ausencia de una dirección definida. Cada vez que se identifica una nueva patología, ésta es saludada alegremente. Desde mi punto de vista, debemos esforzarnos por poner un poco de orden en este caos. La investigación sólo debe centrarse en un número limitado de cuestiones, las más importantes desde un punto económico, tal y como recomiendan Gul y Pensendorfer (2008). Una vez que se hayan comprendido bien sus repercusiones, podremos pasar a tratar otros aspectos que amplíen nuestro modelo de toma de decisiones del individuo.

Agradecimientos

Quiero hacer constar mi gratitud a Xavier Calsamiglia, Itzhak Gilboa, Clara Ponsat y Debraj Ray por sus comentarios a este texto. Soy el único responsable de las opiniones que aquí se sostienen.

Bibliografía

- Acemoglu, D. y J. A. Robinson. *Economic Origins of Dictatorship and Democracy*. Cambridge University Press, 2006.
- Akerlof, G. A. y R. Kranton. «Economics and Identity». *Quarterly Journal of Economics* 115 (2000): 715-753.
- Amiel, Y. y F. A. Cowell. «Measurement of income inequality: experimental test by questionnaire». *Journal of Public Economics* 47 (1996): 3-26.
- Aragonés, E., I. Gilboa, A. Postlewaite y D. Schmeidler. «Fact-Free Learning». *American Economic Review* 95 (2005): 1.355-1.368.
- Arrow, K. J. *Social Choice and Individual Values*. John Wiley and sons, 1951.
- y F. H. Hahn. *General Competitive Analysis*. North Holland, 1983.
- Atkinson, A. B. «On the measurement of inequality». *Journal of Economic Theory* 2 (1973): 244-263.
- Banerjee, A. V. «A Simple Model of Herd Behavior». *Quarterly Journal of Economics* 107 (1992): 797-817.
- Benabou, R. «Workings of a City: Location, Education, and Production». *Quarterly Journal of Economics* 108 (1993): 619-652.
- «Equity and Efficiency in Human Capital Investment: The Local Connection». *Review of Economic Studies*, 62 (1996): 237-264.
- y J. Tirole. «Belief in a Just World and Redistributive Politics». *Quarterly Journal of Economics* 121 (2006): 699-746.
- Besley, T. y S. Coate. «Understanding Welfare Stigma: Taxpayer Resentment and Statistical Discrimination». *Journal of Public Economics* 48 (1992): 165-183.
- Blundell, R., W. K. Newey y T. Persson (eds). *Advances in Economics and Econometrics Theory and Applications* (vol. 2). Cambridge University Press, 2006.
- Bolton, G. E. y A. Ockenfels. «ERC: A Theory of Equity, Reciprocity, and Competition». *American Economic Review* 90 (2000): 166-193.
- Calvo-Armengol, A. y Y. Ioannides. «Social Networks in Labor Markets», en L. Blume y S. Durlauf (eds.). *The New Palgrave Dictionary of Economics*. Palgrave MacMillan Press, 2005.
- Camerer, C. F. y G. Loewenstein. «Behavioral Economics: Past, Present, Future», en Camerer, C., G. Loewenstein y M. Rabin. *Advances in Behavioral Economics*. Princeton University Press, 2003.

- Caplin A. y A. Schotter (eds.). *Foundations of Positive and Normative Economics, Methodologies of Modern Economics*, vol. 1. Oxford University Press (en prensa).
- Cervellati, M., J. Esteban y L. Kranich. «Work Values, Endogenous Sentiments and Redistribution». 2008. Inédito.
- Downs, A. *An Economic Theory of Democracy*. Harpers and Bros, 1957.
- Debreu, G. *Theory of Value*. Yale University Press, 1959.
- Della Vigna, S. «Psychology and Economics: Evidence from the Field». *Journal of Economic Literature* (en prensa).
- Duesenberry, J. S. *Income, Saving and the Theory of Consumer Behavior*. Cambridge, Massachusetts: Harvard University Press, 1949.
- Esteban, J. y D. Ray. «On the Measurement of Polarization». *Econometrica* 62 (1994): 819-852.
- «Conflict and Distribution». *Journal of Economic Theory* 87 (1999): 379-415.
- «On the Salience of Ethnic Conflict». *American Economic Review* 98 (en prensa).
- Falk, A. y U. Fischbacher. «A Theory of Reciprocity». *Games and Economic Behavior* 54 (2006): 293-315.
- Fearon, J. «Rationalist Explanations for War». *International Organization* 49 (1995): 379-414.
- Fehr, E. y K. M. Schmidt. «A Theory of Fairness, Competition and Cooperation». *Quarterly Journal of Economics* 114 (1999): 817-868.
- Fryer, R. y M. O. Jackson. «A Categorical Model of Cognition and Biased Decision Making». *The B.E. Journal of Theoretical Economics*, vol. 8, núm. 1 (contribuciones), artículo 6, 2008.
- Fox, C. R. y A. Tversky. «Ambiguity Aversion and Comparative Ignorance». *Quarterly Journal of Economics* 110 (1995): 585-603.
- Gilboa, I. y D. Schmeidler. *A Theory of Case-Based Decisions*. Cambridge University Press, 2001.
- Gilboa, I., A. Postlewaite y D. Schmeidler. «Probabilities in Economic Modeling». 2007 (en prensa).
- Gilboa, I., F. Maccheroni, M. Marinacci y D. Schmeidler. «Objective and Subjective Rationality in a Multiple Prior Model». 2008 (en prensa).
- Goyal, S. *Connections: An Introduction to the Economics of Networks*. Princeton: University Press, 2007.
- Grossman, G. M. y E. Helpman. *Special Interest Politics*. The MIT Press, 2001.
- Grossman, H. I. «A General Equilibrium Model of Insurrections». *American Economic Review* 81 (1991): 912-921.
- Gul, F. y W. Pesendorfer. «Temptation and Self-Control». *Econometrica* 69 (2001): 1.403-1.435.
- (2008) «The Case for Mindless Economics». En Caplin y Schotter (2008), en prensa.
- Hirshleifer, J. «The Paradox of Power». *Economics and Politics* 3 (1991): 177-200.
- Jackson, M. O. «Social and Economic Networks». Princeton University Press, 2008a.
- «Network Formation». En L. Blume y S. Durlauf (eds.). *The New Palgrave Dictionary of Economics*. Palgrave MacMillan, 2008b.
- Kahneman, D. y A. Tversky. «Prospect Theory: An Analysis of Decision under Risk». *Econometrica* 47 (1979): 263-291.
- Kahneman, D. y A. Tversky. «Choices, values and frames». *American Psychologist* 39 (1984): 341-350.
- Laibson, D. «Golden Eggs and Hyperbolic Discounting». *Quarterly Journal of Economics*, 112 (1997): 443-477.
- Laibson, D. y R. Zeckhauser. «Amos Tversky and the Ascent of Behavioral Economics». *Journal of Risk and Uncertainty*, 16 (1998): 7-47.
- Levine, D. «Modeling Altruism and Spitefulness in Experiments». *Review of Economic Dynamics* 1 (1998): 593-622.
- Levitt, S. A. D. y J. A. List. «What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?». *Journal of Economic Perspectives* 21 (2007): 153-174.
- Lindbeck, A., S. Nyberg y J. Weibull. «Social Norms and Economic Incentives in the Welfare State». *Quarterly Journal of Economics* 114 (1999): 1-35.
- Lindbeck, A. y S. Nyberg. «Raising Children to Work Hard: Altruism, Work Norms and Social Insurance». *Quarterly Journal of Economics*, 121 (2006): 1.473-1.503.
- Mas-Colell, A. «The Future of General Equilibrium». *Spanish Economic Review* 1 (1999): 207-214.
- M. D. Whinston y J. R. Green. *Microeconomic Theory*. Oxford University Press, 1995.
- Moffitt, R. «An Economic Model of Welfare Stigma». *American Economic Review* 73 (1983): 1.023-1.035.
- Von Neumann, J. y O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1994.
- Persson, T. y G. Tabellini. *Political Economics. Explaining Economic Policy*. MIT Press, 2000.
- *The Economic Effects of Constitution*. MIT Press, 2003.
- Piketty, T. «Social mobility and redistributive politics». *Quarterly Journal of Economics* 110 (1995): 551-584.
- Powell, R. *In the Shadow of Power: States and Strategies in International Politics*. Princeton University Press, 1999.
- Rabin, M. «Incorporating Fairness into Game Theory and Economics». *American Economic Review* 83 (1993): 1.281-1.302.
- Rabin, M. y R. Thaler. «Anomalies: Risk Aversion». *Journal of Economic Perspectives* 15 (2001): 219-232.
- Ray, D. *Development Economics*, Princeton University Press, 1998.
- «Development Economics». En L. Blume y S. Durlauf (eds.) *The New Palgrave Dictionary of Economics*, Palgrave, MacMillan, 2008a.
- «Inevitable Costly Conflict». 2008b (en prensa).
- Rook, L. «An Economic Psychological Approach to Herd Behavior». *Journal of Economic Issues* 40 (2006): 75-95.
- Rubinstein, A. «Discussion of Behavioral Economics». En Blundell et al., 2006.
- Samuelson, P.A. «The Pure Theory of Public Expenditure». *Review of Economics and Statistics*, 36 (1954): 387-389.
- Skaperdas, S. «Cooperation, Conflict, and Power in the Absence of Property Rights». *American Economic Review* 82, 4 (1992): 720-739.
- Stiglitz, J. «Information and the Change in the Paradigm in Economics». *American Economic Review* 92 (2002): 460-501.
- Strotz, R. H. «Myopia and Inconsistency in Dynamic Utility Maximization». *Review of Economic Studies*, 23 (1956): 165-180.
- Tabellini, G. «The Scope of Cooperation: values and incentives». *The Quarterly Journal of Economics*. 2007 (en prensa).
- Tirole, J. *The Theory of Industrial Organization*. MIT Press, 1998.
- Vives, X. *Oligopoly Pricing: Old Ideas and New Tools*. MIT Press, 2001.

innovaciones radicales: la visión de un economista

NATHAN ROSENBERG

¿Por qué no existe una economía de la ciencia?

Me gustaría empezar con una pregunta: ¿por qué no existe todavía una disciplina reconocida llamada «economía de la ciencia»? Después de todo, la economía como disciplina ha mostrado fuertes tendencias imperialistas en las últimas décadas. Ha colonizado con éxito numerosos ámbitos, pero la economía sigue siendo una excepción. Tenemos una economía de la educación, una economía de la salud, una economía de los hábitos electorales, una economía del divorcio y una de la delincuencia. Respecto a esta última, la economía de la delincuencia, ¿resulta que delinquir es rentable, especialmente cuando, como ocurre a menudo, las probabilidades de ser detenido y castigado son escasas! Como indicación de la elevada categoría de esta clase de investigaciones uno de sus representantes más distinguidos (Gary Becker), recibió el Premio Nobel de Economía.

¿Por qué razón entonces no tenemos una economía de la ciencia o, mejor dicho, puesto que se trata de un fenómeno reciente, por qué ha tardado tanto en aparecer? Esta pregunta resulta especialmente pertinente a la vista de lo que durante mucho tiempo hemos sabido sobre la ciencia. Es decir, que la investigación científica general rinde gigantescos beneficios económicos es, desde hace tiempo, poco menos que un artículo de fe.

Al menos existe una respuesta parcial a la pregunta de por qué ha tardado tanto en surgir una economía

de la ciencia: la economía es una disciplina que estudia los principios implicados en la adquisición de un uso eficiente de unos recursos escasos. Pero hablar de eficiencia en el empleo de recursos requiere la capacidad de realizar alguna clase de comparación explícita entre costes y beneficios. Veamos, en realidad sabemos mucho acerca de aceleradores lineales, máquinas de radiaciones de sincrotrón, telescopios espaciales Hubble, la cartografía del genoma humano, etcétera. De hecho hace algunos años el Congreso de Estados Unidos decidió cancelar la construcción de un acelerador de partículas llamado Supercolisionador Superconductor cuando los costes del mismo ascendieron de 11.000 a 12.000 millones de dólares (de hecho, sólo el proyecto costó más de mil millones).

Pero, aunque resulta relativamente sencillo calcular los costes de la investigación científica, resulta extraordinariamente difícil estimar sus beneficios. Y si insistimos en considerar sólo los beneficios estrictamente económicos entonces sería complicado defender algunos de los proyectos de la llamada Ciencia con mayúsculas (¿Qué beneficios económicos tiene, por ejemplo, el telescopio espacial Hubble?).

Ahora bien, es innegable que la historia de la investigación científica en el siglo xx ha generado una gran cantidad de beneficios inesperados. Pero la noción generalizada de que las probabilidades de obtener beneficios inesperados son limitadas a duras penas puede servir de

guía para determinar la inversión pública en ciencia, o la asignación de un presupuesto fijo a un campo concreto de investigación científica de entre los muchos posibles. En pocas palabras, las incertidumbres asociadas a los posibles beneficios de la investigación científica son sencillamente inmensas, y resulta difícil llevar a cabo una aplicación rigurosa de principios económicos en un ámbito donde los réditos de la utilización de recursos son, en esencia, inconmensurables.

De lo dicho hasta ahora se infiere que, para que sea posible pensar de forma práctica sobre la ciencia y la tecnología en la sociedad moderna es necesario reconocer que debemos, inevitablemente, convivir con un cierto grado de incertidumbre. No obstante quisiera insistir en que esta necesidad no va en detrimento de la utilidad del análisis económico, al menos mientras no alimentemos expectativas descabelladas acerca de lo que puede lograrse sólo mediante el razonamiento abstracto. Porque el análisis económico por sí solo no puede proporcionar una solución pulcramente empaquetada a la adopción de medidas relativas a los asuntos extremadamente complejos que nos atañen. Y tampoco deberíamos esperar que lo haga. Pero sí puede convertirse en una valiosa herramienta a la hora de buscar relaciones causa-efecto, de intentar comprender la manera en que las instituciones y los incentivos moldean el comportamiento humano y a la hora de encontrar sentido a un corpus formidable de datos empíricos e históricos que están al alcance de los verdaderos estudiosos, y del que pueden derivarse importantes enseñanzas para diseñar políticas y crear instituciones.

Cambios institucionales en el siglo *xxi*

Si repasamos los últimos cien años y nos preguntamos cuáles han sido los rasgos distintivos que han dominado el ámbito de la actividad económica mi respuesta inmediata sería que el principal fue la aplicación del conocimiento y la metodología científicas a un círculo cada vez más amplio de actividades productivas. Pero esta afirmación en sí misma no resulta demasiado informativa. De hecho, sólo puede servir de plataforma a partir de la cual plantear otras cuestiones de mayor calado: ¿exactamente de qué modo ha desempeñado la ciencia este papel?, ¿Qué aspectos de la práctica científica han desempeñado dicho papel y en qué circunstancias? y ¿cuáles fueron los cambios en la manera de institucionalizar la práctica científica en el siglo *xx* que la hicieron tan diferente del siglo *xix*?

Un factor dominante, por supuesto, fue que en los años posteriores a la Segunda Guerra Mundial los gobiernos de los países industrializados se convirtieron, en mayor o menor medida, en patrocinadores de la investigación científica, sobre todo de la investigación básica. Esto reflejaba en grado considerable el papel crítico que la ciencia, y los científicos, habían desempeñado en el curso y en el desenlace de la guerra y que culminó en el desarrollo a manos del Proyecto Manhattan de la terrible arma que

puso abrupto fin a la guerra en el Pacífico. La Guerra Fría contribuyó a cimentar, de forma más poderosa aún, las contribuciones económicas a gran escala a la ciencia en un grado sin precedentes hasta el momento. Pero también estaban activas otras fuerzas, más silenciosas pero igualmente poderosas.

Quizá venga al caso recordar aquí la tantas veces citada observación de Alfred North Whitehead de que «la gran invención del siglo *xix* fue la de inventar una nueva forma de inventar» (Whitehead 1925). El siglo *xx* por supuesto no sólo heredó, sino que también institucionalizó, ese nuevo método de inventar. Whitehead comprendió que esta manera de inventar pasaba por relacionar el conocimiento científico con el mundo de los artefactos. Pero también comprendió que esta relación no era sencilla de establecer, porque entre un descubrimiento científico y la aparición de un nuevo producto o proceso media por lo general un gran trecho. Aunque esta frase que acabo de citar del libro de Whitehead es muy conocida, la que la sigue no lo es tanto, aunque merecería serlo: «Es un gran error pensar que la idea científica por sí sola constituye el invento, y que sólo resta tomar éste y ponerlo en práctica. Antes de que eso suceda es necesario un intenso periodo de diseño imaginativo. Uno de los elementos de una nueva metodología es descubrir cómo salvar la distancia entre las ideas científicas y el producto final. Es un proceso que implica la superación sistemática de una dificultad tras otra».

Lo que parece importar más que la calidad de la ciencia de un país, tal y como se juzga en los círculos académicos o en los criterios del comité de los Premios Nobel, es el grado hasta el cual las actividades de la comunidad científica de dicho país responden a las necesidades de la sociedad en general. Es de lamentar que sea ésta una cuestión rara vez debatida, y menos aún comprendida. A menudo se ve oscurecida por la retórica de la ciencia académica con su énfasis abrumador en el científico individual. El quid de la cuestión es que, a lo largo del siglo *xx* y con diferentes grados de éxito, las sociedades industriales han creado redes cada vez más densas de conexiones institucionales entre el curso de la investigación científica y las necesidades de un sistema social cada vez más grande.

Dentro del mundo universitario esto incluye un número de disciplinas surgidas a finales del siglo *xix* y en el *xx*, tales como la ingeniería eléctrica, la ingeniería química, la aeronáutica, las ciencias metalúrgicas y la informática. De hecho, aunque no parece haber conciencia generalizada de ello, en los últimos años las inversiones gubernamentales en I+D para las universidades norteamericanas dedicadas a las ingenierías han excedido a las dedicadas a las ciencias físicas. Con mucho, los principales beneficiarios en los últimos años han sido las ciencias biológicas, que han recibido más del 50% del apoyo federal.

Además de la aparición de nuevas disciplinas, la innovación tecnológica clave del siglo *xx* fue, por supuesto, el laboratorio de investigación industrial. Estos laboratorios

supervisaban la investigación de frontera dentro y fuera de la comunidad universitaria, aunque durante muchos años lo que dominó sus contribuciones a la industria fue la aplicación de conceptos y metodologías científicas relativamente elementales. En el transcurso del siglo xx sin embargo, y en especial a partir de la Segunda Guerra Mundial, la investigación en muchos de estos laboratorios fue ganando en complejidad. En 1992 el *Directory of American Research and Technology* contabilizó alrededor de 12.000 centros no gubernamentales activos dedicados a algún tipo de investigación científica «de utilidad comercial». Y, de acuerdo con las cifras publicadas por la National Science Foundation, más del 30% de la investigación básica en Estados Unidos estuvo financiada por la industria privada.

El laboratorio de investigación industrial es esencialmente una innovación institucional en la que la agenda de investigación científica está determinada en gran medida por las tecnologías industriales. La misión de los científicos industriales es mejorar el rendimiento y la fiabilidad de dichas tecnologías, además de, evidentemente, inventar otras nuevas. Así, el laboratorio de investigación industrial ha ido convirtiendo la ciencia en una institución cuyas metas están determinadas cada vez más por las fuerzas económicas y que cada vez está más centrada en la consecución de objetivos económicos. En el último siglo, la ciencia se ha ido incorporando gradualmente al sistema de crecimiento que ha impulsado a las sociedades industriales al progreso.

Ese sistema de crecimiento, en el cual el cambio tecnológico desempeñó un papel central a lo largo de dos siglos, se ve hoy reforzado por una intensa labor investigadora de fuertes componentes públicos y privados que varían según los países en función de su historia, su cultura, sus modelos políticos y sus prioridades sociales (para más detalles, véase el informe de las Naciones Unidas «Industrial Development Organization, Industrial Development Report 2005, Capability building for catching-up, Historical, empirical and policy dimensions», Viena, 2005).

Además de los requisitos institucionales, la explotación satisfactoria del conocimiento científico ha florecido más en aquellos países industrializados que han ofrecido a los innovadores potenciales acceso directo al capital así como incentivos financieros sustanciosos, y han promovido y formado infraestructuras empresariales y de ingeniería adecuadas. Así, la Rusia zarista del siglo xix alumbró numerosos y brillantes científicos e innovadores, pero su presencia tuvo un impacto apenas perceptible en una sociedad que carecía de una infraestructura empresarial, financiera y de ingeniería adecuada. Por otra parte, la ascensión de Estados Unidos a una posición de liderazgo tecnológico en una serie de sectores industriales antes de la Primera Guerra Mundial se produjo en un periodo en el que sus logros en la ciencia en general eran limitados y, con pocas excepciones, sin gran relevancia internacional. En este

sentido el Estados Unidos de finales del siglo xix y principios del xx presenta interesantes similitudes con el Japón de la segunda mitad del siglo xx. Ambos países adquirieron un crecimiento industrial rápido con una base científica modesta gracias a su gran capacidad para importar y explotar tecnologías extranjeras.

Por otra parte, el relativo estancamiento de la economía británica en el último siglo se ha producido a pesar de sus brillantes y continuadas actuaciones en la vanguardia científica. Hasta hace no mucho tiempo la comunidad científica británica seguía recibiendo más premios nobeles per cápita que la de Estados Unidos. Pero, al mismo tiempo, los británicos no fueron capaces de ser competitivos ni siquiera en innovaciones originadas en Gran Bretaña, como el radar, el motor de reacción, la penicilina o la resonancia magnética. Es más, la revolución en el campo de la biología molecular que empezó con el descubrimiento de la estructura de la doble hélice de la molécula del ADN en la década de 1950 fue, en gran medida, un logro británico, concretamente de la Universidad de Cambridge, y sin embargo el papel de las compañías británicas en la emergente industria de la biotecnológica fue secundario, comparado con los varios cientos de empresas de biotecnología que se crearon en Estados Unidos, incluidas aquellas pocas que no tardaron en gozar de cierto éxito comercial.

De todo ello quisiera extraer dos conclusiones. Al repasar el siglo xx, los logros científicos por sí solos, por brillantes que hayan sido, no se tradujeron de forma inmediata en un rendimiento económico mayor. Para ello fueron necesarios unos incentivos y un apoyo institucional fuerte, el más importante de los cuales fueron probablemente las sociedades de capital-riesgo. Y es más, allí donde dichas instituciones e incentivos han estado presentes, incluso una capacidad científica comparativamente modesta ha bastado para generar altos niveles de rendimiento económico.

La ciencia endógena

Ya he mencionado que los cambios institucionales del siglo xx han convertido la ciencia en una actividad más endógena, en el sentido de que allí donde se han producido innovaciones tecnológicas, la ciencia ha reaccionado de manera más directa a las fuerzas económicas. Ahora, sin embargo, debo extenderme un poco más en un aspecto concreto de esta afirmación. Lo que pretendo sugerir es que la agenda de investigación de la ciencia ha estado cada vez más condicionada por la necesidad de mejorar la competencia de las tecnologías ya existentes. Durante el siglo xx la ciencia y la tecnología han pasado a estar estrechamente relacionadas. La ciencia se ha convertido en una influencia cada vez mayor en el mundo de la tecnología, pero la causalidad ha sido mutua: la empresa científica del siglo xx también necesita ser explicada en términos de sus respuestas a las necesidades y las exigencias de la tecnología.

De hecho, un asunto que ha sido ignorado por la ciencia del pasado siglo es que el progreso en el ámbito tecnológico

ha pasado a desempeñar un papel clave en la formulación de una agenda de investigación científica. La trayectoria natural de determinadas mejoras tecnológicas ha servido para identificar y definir los límites de futuras mejoras, lo que a su vez ha servido para hacer lo propio con la investigación científica.

Pensemos por ejemplo en la industria aeronáutica. En ella, las mejoras tecnológicas han hecho posible que el avión alcance niveles de rendimiento cada vez mayores y que sólo puedan analizarse con una mayor comprensión de las leyes de la física. Como resultado de ello, la aparición del turboreactor tuvo un profundo impacto tanto en la ciencia como en la industria aeronáutica, al extender progresivamente los límites de las fronteras científicas y al identificar las direcciones específicas en las que este nuevo conocimiento debía seguir antes de que pudieran llevarse a cabo las mejoras tecnológicas.

Así, el turboreactor hizo posible en primera instancia la creación de una nueva especialidad, la aerodinámica supersónica que, según un experto en la materia «pronto cedió el paso a la aerotermodinámica conforme turbo-reactores cada vez más potentes llevaban a los aviones a reacción a velocidades a las cuales la generación de calor en la superficie del avión pasaba a ser un factor esencial en el comportamiento del flujo de aire. Con el tiempo, un avión propulsado por turboreactor pudo alcanzar velocidades a las cuales las consideraciones magneto-termodinámicas cobrarían una importancia capital. [Es decir], las temperaturas se elevaron tanto que el aire se disociaba en iones submoleculares» (Constant 1990). Así, el aumento de la velocidad que los motores a reacción hicieron posible también requería expandir las fronteras del conocimiento científico para poder adaptarse a los requisitos de diseño del nuevo avión.

Lo que sugiero es que uno de los rasgos centrales de las industrias de alta tecnología es que este tipo de secuencia ha pasado a ser la dominante. Es decir, que el progreso tecnológico sirve para identificar, de formas razonablemente precisas, las direcciones en las que la investigación científica debe orientarse y, al mismo tiempo, hace prever potenciales beneficios económicos en caso de que la investigación tuviera éxito.

Los mecanismos que operan aquí pueden adoptar gran cantidad de formas. En el caso del motor a reacción que pasó a funcionar a velocidades cada vez mayores, la tecnología apuntaba a un fenómeno natural específico en un entorno determinado. En la industria del teléfono, por el contrario, la transmisión entre largas distancias o la introducción de nuevas modalidades de transmisión han resultado ser mecanismos especialmente valiosos a la hora de generar investigación básica. Por ejemplo, para mejorar la transmisión transoceánica por radioteléfono era esencial desarrollar un conocimiento mayor de las formas en que las radiaciones electromagnéticas interactúan con las cambiantes condiciones atmosféricas. De hecho, algunos

de los proyectos de investigación científica más importantes del siglo xx fueron consecuencia directa de los intentos por mejorar la calidad de la transmisión del sonido por teléfono. Estudiar los distintos tipos de interferencias, distorsiones o atenuación de las señales electromagnéticas que transmiten sonido ha aumentado significativamente nuestra comprensión del Universo.

Dos avances científicos fundamentales, uno debido a Karl Jansky a finales de la década de 1920 y otro más reciente, de Penzias y Wilson, fueron el resultado de los intentos por mejorar la calidad de la transmisión telefónica, en concreto por buscar las fuentes de ruido. Jansky recibió el encargo de trabajar en los problemas del ruido estático tras la instalación del primer servicio de radioteléfono transoceánico. Para ello debía ayudarse de antenas de radio rotatorias. En 1932 publicó un artículo identificando tres fuentes de ruido: el procedente de tormentas cercanas, un segundo de tormentas más distantes y una tercera fuente, que él llamó «un siseo constante de origen desconocido». Este «ruido estelar», como se llamó en un principio, marcó el nacimiento de una nueva ciencia: la radioastronomía, una disciplina que se convertiría en una de las grandes fuentes de progreso científico del siglo xx.

La experiencia de Jansky viene a resaltar una de las razones por las que los intentos de distinguir entre investigación científica básica y aplicada son extremadamente difíciles de respetar de manera consistente. Los avances fundamentales a menudo se producen tratando de solucionar problemas de índole práctica. Tratar de dibujar una línea de separación basada en las motivaciones de la persona que realiza la investigación —ya sea adquirir información útil (y por tanto aplicada) o bien una búsqueda totalmente desinteresada de nuevos conocimientos (investigación básica)— es a mi juicio una empresa inútil. Cualesquiera que sean las intenciones a priori a la hora de emprender una investigación, el tipo de conocimiento que se resultará de ésta es altamente impredecible. Tal es la naturaleza de la investigación científica más seria. Históricamente algunos de los más importantes avances científicos han procedido de personas, por ejemplo Jansky, que actuaban convencidas de estar haciendo investigación aplicada.

El gran avance de los laboratorios Bell en astrofísica también estuvo directamente relacionado con los intentos por mejorar las transmisiones telefónicas, y en especial en el uso de los satélites de comunicación para dicho fin. En frecuencias muy altas la lluvia y otras condiciones atmosféricas se convierten en importantes fuentes de interferencia. Esta fuente de pérdida de señal era una preocupación constante en el desarrollo de la comunicación por satélite y propició numerosas investigaciones tanto en los campos de la ciencia básica como la tecnológica, por ejemplo, el estudio de los fenómenos de polarización (Dinn 1997).

Arno Penzias y Robert Wilson observaron por primera vez la radiación cósmica de fondo, hoy considerada confir-

mación de la teoría del *Big Bang* de la formación del universo, en 1964, mientras trataban de identificar y medir las distintas fuentes de ruido en su sistema receptor y en la atmósfera. Encontraron que «la radiación está distribuida de forma isotrópica en el espacio y que su espectro es aquel de un cuerpo oscuro a una temperatura de 3 kelvin» (Fagen 1972). Aunque Penzias y Wilson entonces lo ignoraban, el carácter de esta radiación de fondo era precisamente la que había sido predicha anteriormente por los cosmólogos que habían formulado la teoría del *Big Bang*. Su trascendental hallazgo les valió el Premio Nobel.

Lo que estoy sugiriendo es que existe una fuerte lógica interna en algunas industrias, como la del teléfono, que orienta la investigación en direcciones específicas. Consideremos ahora algunas de las necesidades materiales del teléfono. La invención del transistor y el descubrimiento del efecto transistor fueron el resultado de intentos deliberados por encontrar un sustituto del tubo de vacío en la industria del teléfono. El tubo de vacío era inestable y generaba gran cantidad de calor. Después de la invención del transistor su producción en serie requería estándares de pureza del material que no tenían precedentes en la fabricación con fines industriales. Puesto que el funcionamiento del transistor dependía de la introducción de unos cuantos átomos de impureza (de otra sustancia) en el cristal semiconductor, era necesario obtener altos niveles de pureza en dicho semiconductor. Un solo átomo de otra sustancia por cada 100.000.000 de átomos de germanio significaba que el sistema del teléfono debía alcanzar unos niveles de pureza que hacían necesaria gran cantidad de investigación sobre la estructura y el comportamiento de los materiales, en especial el cristal.

La invención del transistor en 1947 tuvo un enorme impacto en la orientación de la investigación científica. Hasta entonces la física del estado sólido había atraído sólo a un reducido número de estudiosos. De hecho, antes de la Segunda Guerra Mundial era una asignatura que ni siquiera se enseñaba en las universidades norteamericanas. Sin embargo, en los años siguientes al anuncio del efecto transistor la asignación de fondos para la investigación experimentó una profunda transformación. De hecho, en cuestión no ya de décadas, sino de años, la física del estado sólido se había convertido en la rama más importante de la física. La gran movilización de recursos económicos hacia ese campo, tanto en las universidades como en la industria privada, fue consecuencia directa de los enormes beneficios potenciales que la llegada del transistor hacía presagiar.

La expansión de la industria del teléfono también supuso que el equipo y los componentes debían funcionar en condiciones ambientales extremas, desde satélites geosíncronos a cables transatlánticos. Estas condiciones extremas tienen una consecuencia particularmente importante: encierran altas posibilidades de traer consigo elevadas penalizaciones económicas si no se consiguen

establecer los estándares requeridos de fiabilidad. Existen razones de peso para la consecución y el mantenimiento de dichos estándares que no son aplicables, por ejemplo, en la electrónica de consumo, por no hablar de una fábrica de ladrillos. El mal funcionamiento de un cable submarino una vez instalado en el lecho oceánico supone elevadísimos costes de reparación o de sustitución de piezas, además de la sustanciosa pérdida de beneficios. Del mismo modo, los satélites de comunicaciones tienen que ser altamente fiables y resistentes sólo para poder resistir a su lanzamiento y puesta en órbita. La instrumentación debe estar diseñada para soportar choques, vibraciones, cambios de temperatura, radiaciones, etcétera.

Así pues, un alto grado de fiabilidad no es una mera consideración secundaria sino la esencia misma del éxito económico en esta industria, y tuvo mucho que ver con la alta prioridad que los laboratorios Bell concedieron a la investigación de materiales durante varias décadas. Así, dichos laboratorios lograron avances importantes en la química de los polímeros con el fin de comprender la morfología del polietileno, debido al deterioro prematuro del revestimiento hecho de este material en el lecho del océano Atlántico.

La importancia de altos niveles de fiabilidad también ha sido una condición básica a la hora de impulsar la investigación en otras direcciones específicas. La decisión de poner en marcha un programa de investigación básica sobre la física del estado sólido, que culminó con el desarrollo del transistor, estuvo fuertemente influenciada, como ya he apuntado anteriormente, por estas (y otras) fuentes de insatisfacción. Pero en sus primeros años el transistor poseía sus propios problemas de fiabilidad, surgidos a principios de la década de 1950 conforme crecieron sus aplicaciones. Pasado un tiempo se descubrió que los defectos tenían que ver con ciertos fenómenos de superficie y como resultado de ello se dio un gran impulso a la investigación de la física de superficies, la cual terminó por resolver los problemas de fiabilidad pero, al mismo tiempo, generó una gran cantidad de información nueva sobre esta rama de la física.

El desarrollo de la fibra óptica resulta un ejemplo especialmente apropiado en esta exposición. Aunque su atractivo como nueva modalidad de transmisión aumentó por las restricciones de espacio y la sobrecarga en las comunicaciones, su viabilidad llegó de la mano de una serie de avances tecnológicos ocurridos en la década de 1950. Fue el desarrollo de la tecnología láser lo que hizo posible la utilización de la fibra óptica para la transmisión. Esta posibilidad, a su vez, apuntó al campo de la óptica, donde los nuevos descubrimientos hacían presagiar sustanciosos beneficios económicos. Como resultado de ello la óptica como campo de investigación científica ha experimentado un importante auge en las últimas décadas. Debido a un cambio de expectativas basado en innovaciones tecnológicas pasadas y prospectivas, pasó de ser una disciplina relativamente estancada e intelectualmente somnolienta a

un campo emergente de investigación científica. Las causas de ello no fueron inherentes al campo de la óptica sino que estaban basadas en una evaluación radicalmente distinta de sus nuevas posibilidades tecnológicas, las cuales, a su vez, tenían su raíz en el avance tecnológico que supuso la aparición del láser.

Esta argumentación posee implicaciones, me parece, que revisten una importancia fundamental a la hora de comprender el papel de la ciencia en la economía moderna. Aunque el impacto del nuevo conocimiento científico en la industria se ve continuamente subrayado en los debates públicos, la atención prestada a las fuerzas causales que fluyen en dirección opuesta es mínima. Y, sin embargo, las industrias de alta tecnología ponen en funcionamiento fuerzas extremadamente poderosas que estimulan e influyen la investigación científica. Ello ocurre de diversas maneras: al proporcionar observaciones o plantear cuestiones que sólo pueden darse en un contexto científico, tales como el teléfono o la industria aeronáutica; al proveer nuevas técnicas de instrumentación que amplían de forma considerable la capacidad de observación, medida y estimación del científico, y, sobre todo, al elevar los beneficios económicos de los resultados de la investigación científica y, por lo tanto, incrementar poderosamente las motivaciones de la industria privada para financiarla.

Debemos entender que los notables logros de los laboratorios Bell durante el siglo xx no tuvieron parangón en otros sectores de la industria estadounidense; de hecho fueron verdaderamente únicos en muchos aspectos, pero muchas otras compañías de este país desarrollaron potentes capacidades científicas de gran valor económico, una afirmación que se ve confirmada por la formulada anteriormente de que en 1992 en Estados Unidos existían alrededor de 12.000 laboratorios de industrias privadas.

Una generalización justa sería afirmar que las compañías estadounidenses aprendieron a explotar el conocimiento y la metodología científicos y a unir estas fuerzas por medio de la organización y los incentivos apropiados, y en dicha empresa tuvieron más éxito que otros países de la OCDE.

El carácter cada vez más multidisciplinar de la naturaleza de la investigación (y la innovación)

Hay otro rasgo de la práctica científica que requiere atención por la importancia de sus implicaciones en el futuro. La naturaleza cada vez más multidisciplinar de la investigación en los ámbitos tanto de la ciencia como de la tecnología a partir de la segunda mitad del siglo xx.

La historia sugiere que el traspase de fronteras disciplinares no es algo que surja de un plan deliberado, de una estrategia o de la reunión de un comité; más bien se trata de algo que ocurre, cuando lo hace, como consecuencia de la peculiar lógica del progreso científico. Históricamente han surgido problemas en la frontera de una disciplina concreta, como puede ser la biología celular, que requerían una mejor comprensión del papel de deter-

minados procesos que eran la especialidad de científicos de otra disciplina, por ejemplo, la química. La disciplina resultante, la bioquímica, es por tanto el producto natural de las cambiantes exigencias de la investigación científica. De igual modo, la geofísica surgió como una rama independiente de la geología cuando se hizo posible aplicar las metodologías, primero desarrolladas en el campo de la física, para la comprensión de la estructura y la dinámica de la Tierra, así como, con el tiempo, de otros planetas. Aquí, como en otras ocasiones, la introducción de nuevas tecnologías de instrumentación ha conducido a una intersección beneficiosa entre disciplinas científicas. Las fronteras tradicionales entre la física y la química se han traspasado en numerosas ocasiones en el pasado por razones similares.

La importancia creciente de la capacidad de explotar el conocimiento y las metodologías de más de una disciplina se ha hecho evidente no sólo en la ciencia básica sino también en la aplicada y en las ingenierías. En los últimos años la ciencia médica se ha beneficiado inmensamente no sólo de disciplinas «cercanas» como la biología y la química, sino también de la física nuclear (resonancia magnética nuclear, inmunoradiensayo), la electrónica, la ciencia de materiales y la ingeniería. En farmacia ha habido avances notables derivados de campos como la bioquímica, la biología molecular y celular, la inmunología, la neurobiología y la instrumentación científica. Estos avances apuntan a la posibilidad de que nuevos medicamentos con propiedades específicas podrán ser fijados como objetivos e incluso, algún día, diseñados, lo que supone un gran avance comparado con los exhaustivos, costosos y aleatorios métodos de testeo que han caracterizado la investigación farmacéutica en el pasado (ver Gambardella 1995). El nuevo patrón de innovación es, por su naturaleza misma, altamente multidisciplinar. Su éxito pasa por la estrecha colaboración entre un número cada vez mayor de especialistas: químicos, bioquímicos, farmacéuticos e informáticos. Sobre lo que no hay duda es que las ciencias biológicas serán cada vez más relevantes en el descubrimiento y desarrollo de nuevos medicamentos. Esto parece también evidente en la todavía naciente industria de la biotecnología, que se nutre de muchas disciplinas científicas diferentes, incluyendo la biología celular, molecular, la química de proteínas y la bioquímica.

El creciente valor de la investigación interdisciplinar plantea serios problemas organizativos de cara al futuro. Esta clase de investigación a menudo va en contra de las costumbres, formación, prioridades intelectuales y dinámica de incentivos de la profesión científica, especialmente en el terreno académico, donde se otorga una inmensa importancia al trabajo realizado dentro de unos límites disciplinares claramente diferenciados. Las disciplinas divididas en departamentos han desempeñado un papel crucial en la enseñanza y la investigación y ciertamente no deben ser menospreciadas. Históricamente, las

disciplinas científicas han surgido porque dentro de sus fronteras existía una serie de problemas que no podían resolverse mediante una conceptualización, un marco o una metodología analíticas corrientes. Los que trabajan en un mismo campo hablan un lenguaje común y, lo que es igualmente importante, esta disciplina que comparten les proporciona una base para elaborar juicios acerca de la cualidad de su investigación. A este respecto la dedicación a una disciplina en particular aseguraba unos mínimos estándares de control de calidad.

Aunque existe una gran (y comprensible) preocupación por la futura financiación de las universidades, también se están planeando interrogantes acerca de su organización. Una compartimentalización rígida de disciplinas va en contra de una investigación que requiere cada vez más traspasar las fronteras disciplinares tradicionales. Cabe observar que dichos problemas no resultan tan graves en el sector de la industria privada, donde las fronteras interdisciplinares no pesan tanto y donde las prioridades son resolver problemas, mejorar el rendimiento de la tecnología existente y, en última instancia, generar mayores beneficios independientemente de las disciplinas científicas aplicadas.

Persistencia de la incertidumbre

Existe una cuestión de la que debemos ocuparnos: se trata de la persistencia de la incertidumbre, no sólo en el terreno de la ciencia, donde está universalmente aceptada, sino también en el de la tecnología. Estamos acostumbrados a esperar un alto grado de incertidumbre y de desenlaces imprevistos en el mundo de la investigación científica. Estas incertidumbres disminuyen conforme avanzamos en el espectro de actividades que van desde la investigación básica a la aplicada, después al diseño y desarrollo del producto y, por último, a la comercialización del nuevo producto en el mercado.

Es cierto, claro está, que algunas de las incertidumbres son resueltas una vez que la nueva habilidad electrónica ha sido establecida. Pero incluso una vez ha sido aceptada en el mercado los interrogantes cambian y estos nuevos interrogantes no son, en modo alguno, más sencillos que los anteriores. El más importante de ellos es: ¿qué función desempeñará esta nueva capacidad tecnológica dentro de la sociedad?

Parece ser que nadie previó la aparición de Internet y que en lugar de ello simplemente «apareció» una vez que hubo un número suficiente de ordenadores funcionando. Tal y como David Mowery apuntaba en su fascinante artículo: «Internet es la red informática más grande del mundo, un conjunto en constante crecimiento de más de cien millones de ordenadores que se comunican los unos con los otros utilizando una red común de normas y protocolos. Gracias a las innovaciones en *software* que han aumentado la accesibilidad y utilidad de la red, Internet ha impulsado una revolución en las comunicaciones que

ha cambiado la manera en que los individuos y la instituciones usan los ordenadores en una gran variedad de actividades» (Mowery y Simcoe 2002).

Pensemos en el láser, una innovación que es sin duda uno de los avances tecnológicos del siglo xx más poderosos y versátiles y del que probablemente restan por descubrir nuevos usos. Desde que se inventó hace más de medio siglo su amplia gama de aplicaciones ha resultado ser asombrosa. Éstas incluyen medición de precisión, instrumentos de navegación y, por supuesto, investigación farmacéutica. Resulta esencial para la reproducción de alta calidad de música en los discos compactos o CD. Se ha convertido en una alternativa a numerosos procedimientos quirúrgicos, incluida la extremadamente delicada cirugía ocular, donde se emplea para corregir desprendimientos de retina, y en la ginecológica, donde ha hecho posible la extirpación de determinados tumores mediante procedimientos más sencillos y menos dolorosos. Se emplea extensamente también en cirugía de la vesícula biliar. Las páginas de este artículo han sido impresas en una impresora láser. El láser se emplea diariamente en la industria, incluida la textil, donde permite cortar las telas en las formas deseadas, en metalurgia y en la fabricación de materiales compuestos, donde desempeña funciones similares. Pero quizá donde mayor impacto ha tenido la aplicación del láser es en el mundo de las telecomunicaciones donde, junto con la fibra óptica, está revolucionando la transmisión. El mejor cable trasatlántico existente en 1996 era capaz de transportar simultáneamente sólo 138 conversaciones entre Europa y América del Norte. El primer cable de fibra óptica, instalado en 1988, podía transmitir hasta 40.000. Los cables de fibra óptica instalados a principios de la década de 1990 transmiten casi un millón y medio de conversaciones al mismo tiempo. Y sin embargo sabemos que los abogados de patentes de los laboratorios Bell se mostraron reacios en un principio a solicitar la patente del láser argumentando que tal invento, puesto que entraba en el terreno de la óptica, no tendría aplicación posible en la industria del teléfono. En palabras de Charles Townes, quien recibió el Premio Nobel de Física por sus investigaciones con el láser, «el departamento de patentes de Bell rechazó en primera instancia patentar nuestros amplificador y oscilador para frecuencias ópticas porque, aducían, las ondas ópticas nunca habían sido relevantes para las comunicaciones y por tanto nuestro invento no tenía interés alguno para los laboratorios Bell» (Townes 1968).

El transistor fue, sin duda, uno de los grandes avances del siglo xx, incluso, podríamos decir, de la historia de la humanidad. En consecuencia cabría esperar que el anuncio de su invención, en diciembre de 1947, hubiera ocupado la primera página del *New York Times*. Nada más lejos de la realidad. Cuando por fin se mencionaba, era sólo en un breve en las páginas interiores, dentro de una columna semanal titulada «Noticias de la radio». Aparte de la fabricación de mejores audífonos, no se le suponía ninguna otra utilidad práctica.

Si hiciéramos una relación de grandes inventos científicos del siglo xx cuyas aplicaciones futuras y beneficios potenciales pasaron desapercibidos en su momento sería infinita: el láser, el ordenador, el transistor... Si quisiéramos, podríamos divertirnos recordando la incapacidad de generaciones anteriores para ver lo obvio, tal y como lo vemos hoy día. Pero eso sería, a mi juicio, una presunción por nuestra parte. No tengo especial fe en nuestra capacidad para superar las incertidumbres que van aparejadas al uso de nuevas tecnologías.

De igual manera, la principal razón para las modestas perspectivas de futuro que se perfilaban para el ordenador a finales de la década de 1940 era que los transistores no se habían incorporado aún a dichos aparatos. Al introducir el transistor, y más tarde los circuitos integrados, la industria informática se transformó por completo. De hecho, en uno de los más extraordinarios logros del siglo xx, el circuito integrado con el tiempo se *convirtió* en un ordenador, gracias a la aparición del microprocesador, en 1970. El mundo sería hoy muy distinto si los ordenadores siguieran operando con tubos de vacío.

Bibliografía

- Constant, E. W. *The Origins of the Turbojet Revolution*. John Hopkins University Press: Baltimore, 1990: 240.
- Dinn, N. F. «Preparing for future Satellite Systems». *Bell Laboratories Record* (octubre de 1977): 236-242.
- Fagen, M. D., ed. *Impact. Bell Telephone Laboratories*. 1972: 87.
- Gambardella, A. *Science and Innovation in the US Pharmaceutical Industry*. Cambridge University Press, 1995.
- Mowery, D. y T. Simcoe. «Is the Internet a US Invention?». *Research Policy*, vol. 31, (diciembre de 2002): 1.369-1.387.
- Townes, Ch. «Quantum Mechanics and Surprise in the Development of Technology». *Science* (16 de febrero de 1968): 701.
- Whitehead, A. N. *Science and the Modern World*. The MacMillan, 1995: 98.

por qué es tan difícil combatir la pobreza

ABHIJIT V. BANERJEE

Uno de los motivos por los que las medidas para combatir la pobreza no han funcionado mejor es porque hemos abordado el tema de forma ingenua, sin comprender bien dónde reside su complejidad.¹ Este artículo trata de lo que he averiguado sobre este asunto en mis investigaciones, basadas en su mayoría en la India.

Identificar a los pobres

¿Quiénes son los pobres?

Supongamos que alguien quiere ayudar a los pobres, ¿cómo los encuentra? Una parte del problema es inevitable: «pobre» es una categoría relativa como pueden serlo «alto» o «guapo». Si bien por lo general sabemos de qué estamos hablando cuando nos referimos a pobres, obtener una definición funcional de la pobreza implica una serie de elecciones bastante arbitrarias. Por ejemplo, incluso si estamos dispuestos a arriesgarnos y señalar que las personas que están por debajo de un determinado nivel (la llamada línea de pobreza) son pobres y los demás no, antes deberemos precisar qué nivel es ése y qué factores definen su línea divisoria. Los más obvios serían la renta, el consumo y la riqueza, pero sin duda podríamos pensar en otros. De todos ellos, la renta parece la más lógica, hasta que uno empieza a pensar en las dificultades que supone medirla: después de todo la renta varía mucho, sobre todo para los pobres que no tienen un trabajo asalariado, y pueden

tener ingresos distintos de un día o de un mes a otro, en ocasiones fortuitamente o de manera deliberada (pensemos en el vendedor que se toma un día libre por semana) y que no repercuten en lo que pueden comprar o consumir (porque gastan de sus ahorros o toman prestado). En otras palabras, corremos el riesgo de concluir que el vendedor es pobre porque hemos medido los ingresos de su día libre.

Calcular la renta media durante periodos de tiempo más prolongados nos sirve de ayuda en este aspecto, pero suscita otra clase de problemas. A la gente no se le da bien acordarse de lo que ocurrió hace unas semanas o unos meses, sobre todo si existen muchas variaciones posibles. También le resulta muy difícil calcular su propia renta (a menos que ganen un salario, e incluso en estos casos puede que desconozcan el valor de las prestaciones que incluye el puesto de trabajo). Ello se debe en parte a que tienen tanto entradas como salidas de dinero (es decir, ganancias y gastos), y éstas no se producen al mismo tiempo (con lo que hay que buscar la manera de compararlas).

Por estos motivos muchos economistas prefieren utilizar como medida el consumo, que sin duda varía mucho menos que los ingresos (lo que pone de manifiesto que los individuos tienden a evitar modificar sustancialmente sus hábitos de consumo) y, por consiguiente está estrechamente relacionado con la renta media de un periodo de tiempo determinado. Esta medida también presenta limitaciones: de manera sistemática subestimamos el bienestar de los que ahorran mucho en comparación con los que

¹ Los razonamientos que apoyan esta tesis están expuestos en Banerjee (2007).

no lo hacen, a pesar de que estos últimos tienen mejores perspectivas de futuro. La cuestión del gasto en cuidados sanitarios plantea además un problema añadido: ¿debemos excluir estos gastos cuando calculamos el consumo basándonos en que son obligatorios y no una elección, o incluirlos en tanto que nos indican si una familia determinada es capaz de hacerles frente y otra más pobre tendrá que resignarse en cambio a una mala salud?

Así que medir el consumo es probablemente más fácil que medir la renta (sobre todo porque los individuos tienden a adoptar patrones de consumo relativamente estables, y por tanto podemos alcanzar una conclusión razonable preguntándoles cómo han gastado el dinero en los últimos tiempos), aunque ello tampoco es una tarea sencilla. Para empezar, puede llevar muchísimo tiempo: a los individuos les cuesta trabajo recordar lo que han consumido la semana anterior, a no ser que les presenten una lista completa y específica de los productos que hayan podido comprar y les pregunten por cada uno de ellos por separado. Asimismo, las decisiones relativas al consumo varían en función del sexo: por lo general, los hombres saben mucho más del dinero que gastan en los arreglos de la casa, mientras que las mujeres tienen una idea mucho más precisa de lo que cuestan las cebollas. Por este motivo puede que sea necesario preguntar a más de una persona dentro de un mismo hogar para hacerse una idea precisa de sus gastos de consumo.

El procedimiento de identificación

Dado el enorme tiempo que lleva y lo meticuloso que hay que ser a fin de obtener un indicador exacto de la pobreza basado en la renta o el consumo, quizás no nos sorprenda que la mayoría de los gobiernos de los países en vías de desarrollo opten por estimaciones más aproximadas y un enfoque simplificado. En lugar de buscar indicadores directos relacionados con el consumo o la renta, utilizan unos test por aproximación de recursos llamados *proxy means tests*. En ellos cada hogar recibe una puntuación basada en un número relativamente pequeño de lo que se consideran indicadores representativos del nivel de vida de una familia. Para determinar qué segmento de población de la India vive por debajo del nivel de pobreza, se emplea un sistema de puntuación que tiene en cuenta la riqueza de la familia (ser propietario de tierras, el tipo de casa, si ésta tiene fontanería, etcétera); el grado de bienestar (como, por ejemplo, si toman dos comidas decentes al día); aspectos relativos a la capacidad de ganar dinero (el nivel educativo de los adultos, el puesto de trabajo que tienen, etcétera), y algunos índices que podrían denominarse «reacciones conductistas» a la pobreza (si los niños están escolarizados, trabajan, etcétera). El programa insignia de bienestar de México, que en la actualidad se llama Oportunidades, emplea un índice parecido para identificar a los posibles beneficiarios de prestaciones que consiste en una media ponderada del número de personas por habitación en cada

hogar, la edad del cabeza de familia, la tasa de dependencia, el nivel educativo y la profesión del cabeza de familia, el número de hijos entre 5 y 15 años sin escolarizar, el número de hijos menores de 12 años y algunas variables binarias simples que sirven para definir el tipo de vivienda así como el patrimonio de cada familia. Los diversos programas de asistencia focalizados que existen en Indonesia emplean un sistema parecido, si bien algo más sofisticado.

La ventaja que tiene usar un sistema de estas características es que la información necesaria se puede recopilar en media hora o menos; el inconveniente es que puede que no siempre nos conduzca adonde queremos llegar. Al utilizar datos de Indonesia, Nepal y Paquistán en los que se recogía información sobre el consumo y el patrimonio, Filmer y Pritchett (2001) señalan que sólo entre el 60 y el 65% de los ciudadanos que entraban dentro del 40% de la distribución inferior, tomando como base el consumo, entraban también en el 40% inferior de la que se había hecho tomando como base el patrimonio. En otras palabras, alrededor de un 35-40% de los pobres podrían no estar correctamente clasificados, si bien es probable que sean menos, puesto que no hay ninguna razón para suponer que los indicadores de consumo sean siempre fiables.

No obstante, hay otra cuestión preocupante. Emplear formas determinadas de riqueza como indicadores tiene la ventaja de que es muy fácil de medir, pero el inconveniente de que también es muy fácil de manipular: si creo que construir otra habitación en mi casa disminuirá mis posibilidades de recibir ayuda del gobierno, puede que decida invertir mis ahorros en oro. Esto se vuelve aún más peligroso cuando la elección se basa en si un hijo debe ir o no a la escuela, pues puede que los padres, que todavía no estén convencidos de los beneficios de recibir una educación (más tarde trataré este asunto), no se lo piensen mucho antes de sacar a sus hijos del colegio a fin de asegurarse un puesto en la lista de beneficiarios de ayudas del Estado.

La dificultad de la puesta en práctica

Sin duda alguna, todo método de identificación de los verdaderamente pobres será sólo lo bueno que permitan las personas que lo emplean. Como ya hemos señalado, determinar quiénes son los pobres supone una gran dificultad, incluso en los casos en que los que se emplean criterios simplificados, y no está claro que los responsables tengan razones de peso para hacerlo bien. En efecto, no es difícil imaginar que es posible que la persona que decide si usted puede entrar o no en la lista de los que necesitan ayuda del Estado quiera cobrar algo por hacerle el favor, y, si usted es realmente pobre y no puede permitirse lo que le pide, tal vez prefiera darle su tarjeta a otro que se lo merezca menos pero que le pueda pagar. Además, hay una tendencia natural a mostrarse generoso a la hora de interpretar las reglas: ¿por qué privar a alguien de ayuda simplemente porque no se ajusta a

los criterios, cuando el riesgo de que alguien se queje es prácticamente inexistente?

En esta misma línea, un estudio reciente realizado en la India en el que se compara el número de personas pobres con el número de tarjetas BPL (siglas de Beyond the Poverty Level, por debajo del umbral de pobreza) que se han emitido, llegó a la conclusión de que había 23 millones de titulares más de los que debería (NCAER 2007, información sacada del *Times of India*, 22 del 12 de 2007). Otro estudio realizado por la ONG Transparency International en colaboración con el Center for Media Studies en la India se centró de manera más directa en la asignación incorrecta de ayudas. Encuestaron a un grupo de familias al azar y les preguntaron acerca de su situación económica y si disponían de una tarjeta BPL (TI-CMS 2007). El estudio puso de manifiesto que dos tercios de las familias que vivían por debajo del umbral de pobreza tenían la tarjeta, lo cual no está muy mal teniendo en cuenta que los indicadores empleados para determinar su nivel económico eran relativamente rudimentarios, y, aun con eso, se superó al estudio de Filmer-Pritchett relativo a la asignación mediante el uso de información relacionada con la riqueza mencionado anteriormente. Por supuesto, también se producen errores de inclusión (de ahí los 23 millones de tarjetas de más), pero esto podría ser simplemente un reflejo del hecho de que es difícil, y, quizás, inútil establecer distinciones precisas dentro de un grupo que en general es pobre.

No obstante, un trabajo más minucioso debido a Karnataka contradice esta interpretación más benévola de la situación. En el estudio de Atanassova, Bertand y Mullainathan (2007) se encuestó a 21 familias en cada uno de los 173 pueblos del distrito de Raichur, en el estado de Karnataka. Recopilaron de cada una de ellas los datos que el gobierno utiliza para hacer la clasificación BPL y los tomaron como base para elaborar su propia lista. Descubrieron que, aunque el 57% de las familias de los pueblos en cuestión tenían tarjetas BPL, sólo un 22% reunía los requisitos necesarios. Además, el 48% no estaba dentro de la categoría correcta. Los errores de inclusión, esto es, familias que cumplían con los requisitos pero que no tenían tarjetas BPL, rondaba el 7%. Esto quiere decir que un tercio de los hogares que cumplían los requisitos no tenía tarjeta, mientras que la mitad de los que no los cumplían disponía de ellas. Y aún hay algo más alarmante, cuando se utilizaba la renta como indicador de riqueza, resultaba que los más pobres de entre todas las familias que no reunían los requisitos no tenían tarjetas BPL. Peor aún, era menos probable que la lista de quienes reunían los requisitos incluyera a aquellos justo por encima del límite de la pobreza, esto es, a los que tenían una renta de entre 12.000 y 20.000 rupias, que a aquellos cuya renta estaba entre las 20.000 y las 25.800 rupias. Además, un 42% de los más ricos (con una renta superior a las 38.000 rupias) tenía tarjeta. Cuando se investigó por qué estas familias que no reunían los requisitos tenían tarjeta,

resultó que estar relacionado socialmente con los funcionarios del pueblo era un factor determinante.

Un enfoque más participativo

El hecho de que el proceso mediante el que se establece quién es pobre esté en manos de la élite del pueblo puede ser uno de los motivos por el que otros han propuesto un enfoque muy diferente: ¿por qué no aprovechar el hecho de que las comunidades pequeñas (pueblos) son capaces de identificar a los realmente pobres dentro de ellas? Y aunque tomados por separado, sus habitantes tengan razones para presentar la información de alguna forma determinada, esto podría evitarse juntando a varios de ellos en un grupo lo suficientemente grande.

Bandhan, una de las instituciones microfinancieras más grandes de la India, aplicó este enfoque para determinar quiénes serían los beneficiarios de su programa Ultra-poor (ultra pobres). En él, a las familias que eran demasiado pobres para cobijarlas bajo el paraguas del microcrédito se les «regalaba» un bien (que podía ser una vaca, unas cuantas cabras o una trilladora) así como un subsidio económico a corto plazo (hasta que el bien que se les había entregado empezara a producir dinero) con la esperanza de que esto pudiera sacarles de su pobreza extrema y permanente y situarles dentro del grupo mayoritario de pobres del pueblo. Siguiendo la metodología que elaboró la ONG bangladesí BRAC, creadora en primera instancia de este programa para identificar los llamados ultra pobres, Bandhan realizó una serie de evaluaciones rurales participativas (Participatory Rural Appraisal, PRA) en el pueblo.² En ellas, un mínimo de 12 vecinos del pueblo, idealmente procedentes de distintos sectores de la comunidad, se sentaban y hacían una planificación en la que a cada familia se le asignaba un hogar y a continuación las dividían en seis categorías, de la más rica a la más pobre. Tras estas evaluaciones, Bandhan seleccionaba a unas 30 familias del grupo de las más pobres.

El proceso de Bandhan no se detuvo aquí. A continuación recopilaban bienes en propiedad y otra información sobre estos 30 hogares y finalmente seleccionaron a 10 de ellos para que formaran parte del programa Ultra-poor. Lo que a nosotros nos interesaba era si este tipo de enfoque resultaba eficaz a la hora de identificar a los pobres y, en algunos aspectos, los resultados ponen de manifiesto que así es (véase Banerjee, Chattopdhyay, Duflo y Saphiro 2008). Los que quedaron clasificados en las dos categorías inferiores en estas evaluaciones contaban con unas 0,05 hectáreas menos de tierra que el resto de la población objeto de estudio, lo cual puede no parecer mucho hasta que se tiene en cuenta que la media de la tierra en propiedad de esta población es en realidad de 0,04 hectáreas. Asimismo, mientras que el 34% de los habitantes encuestados manifestaron que no siempre tomaban una comida decente al día, esta fracción supone otros 17 puntos porcentuales, es decir, es un 50% más elevada entre

2

Las evaluaciones de recursos participativas son una técnica habitual para hacer que un grupo de habitantes planifique su pueblo de manera conjunta.

estos dos grupos más pobres. Estas familias tienen también menos probabilidades de estar escolarizadas y más de tener un hijo que tampoco lo esté así como de contar con una persona discapacitada a su cargo.

Las PRA no nos ayudan pues a determinar quiénes son pobres en relación con el consumo, pero es que también encontramos que, en estos pueblos, poseer una tarjeta BPL tampoco tiene relación con el consumo. Además, a diferencia de dichas tarjetas, las evaluaciones PRA sí pueden pronosticar la escasez de tierra de una familia y el hecho de que no pueda permitirse dos comidas diarias.

Por consiguiente, los habitantes de los pueblos tienen información que pueden y están dispuestos a utilizar para bien del interés público. Incluso más, es posible que la información de la que disponen les permita establecer grados dentro de la población pobre.

Desgraciadamente, al menos en estos pueblos las evaluaciones pasaron por alto a un cuarto de los sujetos que sí figuran en nuestro estudio; sus nombres nunca aparecieron. Y puesto que nuestra labor se centraba de manera específica en los pobres, no es que estas personas no fueran relevantes al objeto de estudio. Básicamente, parece que hasta en los pueblos de pocos cientos de habitantes «no estar a la vista» equivale a «no existir». Las PRA pueden clasificar relativamente bien a quienes encuentran, pero ¿qué ocurre con los que se quedan fuera?

Otro aspecto preocupante del enfoque de estas evaluaciones es que es posible que funcionen mejor a la hora de identificar a los ultra pobres, que al pobre medio. Es probable que la mayoría de los individuos consideren que están por encima de los ultra pobres y por consiguiente sientan una cierta obligación moral de ayudar a éstos. Por otra parte, cuando se trata de determinar quién es el pobre medio, es posible que cada vecino del pueblo piense que se lo merece tanto como cualquier otro, lo que puede dar lugar a desacuerdos y disputas.

En cualquier caso, los resultados de este pequeño proyecto piloto fueron lo suficientemente prometedores como para animarnos a investigar esta cuestión en mayor profundidad. Quizás se deberían combinar ambos enfoques: se podría empezar con una lista de los posibles pobres basada en los datos relativos a la riqueza (y otros aspectos) y, entonces dejar que la comunidad del pueblo la edite (con lo que se reduciría el riesgo de que algunos queden fuera) teniendo en cuenta que disponen de una información más completa. Podríamos plantear otros muchos híbridos también. En un proyecto de investigación todavía en curso, Rema Hanna, Ben Olken, Julia Tobias y yo, del Abdul Latif Jameel Poverty Action Lab (Laboratorio de acción contra la pobreza Abdul Latif Jameel) del Instituto Tecnológico de Massachusetts, en colaboración con el gobierno de Indonesia y Vivi Alatas y su equipo del Banco Mundial en Jakarta, hemos diseñado experimentos para comparar de manera rigurosa la eficacia de este estudio y la metodología empleada para realizar las PRA a

la hora de identificar a los pobres, así como para examinar algunos de estos híbridos.

Autodesignación

La alternativa a la identificación de un objetivo es la auto-identificación o autodesignación. Esta idea, por supuesto, no es nueva. Los famosos asilos para pobres de la época victoriana, que Scrooge elogiaba y sobre los que el caballero compasivo del *Cuento de Navidad* dijo: «Muchos no pueden ir allí; y otros muchos preferirían morir antes que hacerlo», eran exactamente eso: un lugar tan miserable que sólo los terriblemente pobres querrían ir ahí. La India ha introducido hace poco el National Rural Employment Guarantee Scheme (Plan nacional de garantía de empleo rural, NREGS), según el cual todos los hogares rurales tienen derecho a 100 días de empleo público no cualificado a cambio del salario mínimo cuando lo soliciten (con 15 días de antelación) en su pueblo, el mayor esfuerzo realizado en este sentido hasta la fecha.

La teoría en la que se basan estos planes es bien conocida: no es necesario hacer una selección, pues sólo quienes no tienen ninguna alternativa mejor estarán dispuestos a realizar el tipo de trabajos (cavar zanjas, cargar ladrillos) que se ofrecen. Lo cierto es que como se trata de trabajo a demanda, no requiere pedir el permiso de nadie para buscar empleo. Además, tiene la ventaja de que es flexible: gran parte de la pobreza extrema es temporal y/o impredecible. Por ejemplo, cuando la persona que lleva dinero a un hogar cae enfermo, es posible que transcurra mucho tiempo antes de que se clasifique a la familia como BPL, pero el derecho a trabajar está ahí si se quiere ejercer.

Los inconvenientes también son muy obvios: ¿qué ocurre si no hay nadie en la familia en condiciones de realizar trabajo manual? Además, el trabajo es un recurso social: hacer que la gente cave una zanja a fin de demostrar que son pobres es, en efecto, una pérdida de tiempo a no ser que, de hecho, sea necesario cavar esa zanja en particular. Si en cambio la zanja no era necesaria en absoluto y existe otra manera de identificar a los pobres, podíamos haberles dado el dinero y dejarles que hicieran algo productivo con su tiempo. Una parte considerable de los documentos de los NREGS estaba dedicada, por consiguiente, a explicar detenidamente las necesidades de cada pueblo con objeto de asegurarse de que los trabajos realizados fueran de utilidad pública.

La corrupción supone un obstáculo añadido. Desde luego siempre ha sido un problema, pero el hecho de que los NREGS están en función de las necesidades y por tanto no cuentan con un presupuesto fijo hace especialmente tentador incluir unos cuantos nombres de más. Éste es el problema de los registros falsos (donde se recogen todas las transacciones de los NREGS) que han señalado los detractores del programa. Por este motivo se exige que dichos registros estén expuestos en un lugar público. Además, los impulsores del plan insistieron en la necesidad de realizar

lo que denominaron auditorías sociales, mediante las cuales un grupo de voluntarios interesados intenta encontrar a las personas que aparecen en dichos registros y comprueban si han recibido o no los pagos.

Estas auditorías sacaron a la luz bastantes casos de corrupción en la puesta en práctica de los NREGS: en el estado de Jharkhand una auditoría social realizada en cinco pueblos escogidos al azar por los encargados del proyecto de investigación de la Universidad de Allahabad reveló que alrededor de un tercio del dinero había desaparecido (Dreze, Khera y Siddhartha 2008). Lo que es más terrible aún, uno de los activistas que participaron en esta auditoría fue asesinado, se cree que por causas relacionadas con la investigación. Por otro lado, en Chattisgarh, una auditoría de nueve proyectos elegidos también al azar indicó que el 95% de los salarios que figuraban como pagados se cobraron de verdad.

Aunque un 5% de corrupción no está mal y un tercio no está tan bien, sigue sin estar claro cuál debe ser la cuota de referencia. Otra de las críticas al programa es que no está haciendo lo suficiente. El Comptroller and Accounts General (interventor general) de la India, una organización gubernamental responsable de supervisar los programas públicos, reveló que sólo un 3,2% de las personas inscritas habían trabajado el total de 100 días, y que, de media, las familias inscritas habían trabajado menos de 20 días. En respuesta a esto, el Ministerio para el Desarrollo Rural, encargado de la dirección de este programa, señaló que entre las familias que participaban en él (es decir, las que de hecho consiguieron un empleo) la media de los días trabajados estaba más cerca del 40%, y que un 10% habían trabajado los 100 días.

Ahora bien, ¿cómo podemos saber si 40 (o un 10%) constituye demasiado o demasiado poco? Si nadie termina aceptando estos empleos, pero el solo hecho de que existan a un salario mínimo hace subir los salarios del sector privado, donde además no se pierden puestos de trabajo, podría considerarse que el programa es un éxito. También opinaríamos esto si casi nadie pide estos trabajos, pero la simple garantía de que estén disponibles en caso necesario hace que la población esté menos preocupada o más dispuesta a la hora de asumir riesgos económicos. Por el contrario, si todo el mundo quiere uno de estos empleos, pero solo un 50% lo consigue durante el total de 100 días anuales, es probable que nos sintiéramos decepcionados. El informe CAG antes mencionado indica que al menos existe una demanda que no se ha visto satisfecha, y responsabiliza de ello al hecho de que en el programa falta personal, aunque no especifica hasta qué punto.

En el estudio citado anteriormente que llevamos a cabo en el oeste, también encontramos que al menos en los pueblos sujetos de estudio, tener una tarjeta laboral (que es lo que a uno le dan cuando se inscribe en el programa) no implicaba ser pobre. Por consiguiente, cabe preguntarse si esto significa que dicho programa no está

identificando los objetivos debidamente, o si todo el mundo quiere conseguir una de estas tarjetas para tener más seguridad, pero en realidad sólo tienen intención de usarla si no les queda otra alternativa.

Lo que es más importante, incluso si se están identificando los objetivos correctamente y los fallos del programa no son mayores que los de otro, ¿cómo sabemos que merece la pena que la gente dé tantas vueltas para conseguir dinero? En otras palabras, a menos que estemos suficientemente seguros de que los resultados que se obtengan por medio del programa laboral compensen el tiempo y el esfuerzo invertidos, ¿cómo podemos tener la certeza de que tiene sentido pasar por todo el proceso en lugar de tratar de mejorar la identificación de objetivos?

Muchas de estas preguntas habrían tenido respuesta si se hubiera hecho una evaluación rigurosa del programa (junto con un estudio detenido de los diversos grupos que al final no participaron en los NREGS). Sin embargo, la decisión reciente de ampliarlo a todo el país significa que no será así, al menos en la India,³ con lo que la cuestión de si la autodesignación merece la pena queda sin resolver.

El rendimiento de los programas focalizados

El mayor programa focalizado de la India es el Targeted Public Distribution Scheme o TPDS, bajo el cual las familias BPL pueden comprar cereales y otros productos básicos en las denominadas tiendas de precio justo de sus pueblos, las cuales se abastecen de los almacenes del gobierno que haya en la zona. Es el programa que recientemente el ministro de Economía de la India describió en los siguientes términos: «Alrededor del 58% del cereal subvencionado no llega al grupo de personas al que está dirigido, y de éste, algo más del 36% desaparece de la cadena de abastecimiento. Permítanme que les pregunte si los pobres de la India no se merecen un sistema mejor de distribución estatal. ¿Cómo podemos quedarnos de brazos cruzados mientras les es robada la escasa ayuda social que les corresponde?».

Lo más llamativo de estas cifras (sacadas de un informe reciente realizado por la Programme Evaluation Organization (Organización de evaluación de programas) del propio gobierno es que la mayor parte de estas pérdidas no se debe a una asignación errónea de las tarjetas BPL de la que hablamos antes; se trata del robo directo de los cereales en algún momento de la cadena de distribución. De este 36%, un 20% se pierde por el camino de forma misteriosa, mientras que el otro 16% se entrega a tarjetas BPL «fantasma», es decir, emitidas a personas que no existen.

El informe también informa sobre lo que denomina «error de exclusión»: según sus cifras, sólo el 57% de las familias que se encuentran por debajo del umbral de pobreza se benefician de los TPDS. En otras palabras, ni siquiera podemos decir que estas enormes pérdidas sean el precio a pagar por ayudar a los pobres.

3
 Todavía tiene sentido intentar evaluar el efecto que pudiera tener modificar algunos detalles del programa antes de darlo por cerrado. ¿Habrá una mayor creación de empleo si, por ejemplo, en lugar de dar por hecho que todos los beneficiarios del programa tienen que trabajar para la comunidad, se enviase a algunos a trabajar en negocios privados, aunque el gobierno siga apoyando su salario y se asegure de que no perciben menos del salario mínimo? ¿Se impulsaría más el programa si las élites locales tuvieran la impresión de que también les beneficia a ellas? ¿O las animaría a intentar «apropiarse» de él?

Si bien, como hemos señalado antes, la identificación de objetivos es difícil, cuesta creer que el gobierno no pueda hacer algo más por impedir los robos, si es que existe una voluntad política real. En efecto, se puede ver que en dos estados de la India, Tamil Nadu y Bengala occidental, el porcentaje de robos es inferior al 20%.

Pero, puesto que una gran parte del problema es esta falta de voluntad política así como la escasa eficiencia del sistema de identificación de objetivos, puede que sea mejor olvidar esta opción. Con ello eliminaríamos el error de exclusión y atraeríamos a los no pobres, que son los que tienen una mayor influencia en el sistema político, al ámbito del programa.

Ayudarles para que se ayuden a sí mismos

El punto de vista convencional es que el gobierno consigue este objetivo básicamente proporcionando a los hijos de los pobres la atención y la educación necesarias para que, con el tiempo, puedan participar de manera completa en la economía. También puede dar atención sanitaria a adultos como un medio de protegerles frente a posibles contingencias.

La nutrición

La India tiene, con diferencia, el mayor índice de atrofia y raquitismo infantil del mundo. Según el último Estudio nacional de la salud de la familia (NFHS-3 por sus siglas en inglés), el 48% de los niños padece raquitismo y el 43% algún tipo de atrofia, lo que significa que la tasa de estas dos enfermedades en la India es aproximadamente el doble que la de África subsahariana, teniendo en cuenta que se trata de un país mucho más rico.

No obstante, mientras que la desnutrición constituye claramente un enorme problema en la India, no está claro hasta qué punto está relacionada con las prácticas alimentarias antes que con el acceso a la comida. Resulta que estas tasas alarmantes de atrofia y raquitismo corresponden al grupo demográfico medio de los cinco que recoge el NFHS. Cuesta creer que este grupo no pueda permitirse los niveles de nutrición infantil de una familia tipo de un país medio del África subsahariana.

Además, no es evidente que los TPDS, con su diseño actual, supongan una gran aportación a la hora de solucionar los problemas de malnutrición. En parte se trata solamente de una transferencia de ingresos, y los datos obtenidos indican que recibir dinero extra no se traduce en una mejor alimentación (Strauss y Thomas 1998).

El hecho de que estos ingresos adicionales lleguen en forma de alimentos puede que ayude, pero sólo si los 20 kilos de cereales que la familia recibe de los TPDS suponen una cantidad superior a la que habría podido comprar, lo que, se mire como se mire, parece altamente improbable.

Teniendo todo esto en cuenta y a la vista de los desastrosos resultados, puede que la solución de comida sub-

vencionada deba ser replanteada. ¿Por qué no dar dinero en lugar de comida, y evitar así todos los problemas derivados del comercio justo? Es verdad que el precio de los alimentos varía, pero la cantidad de dinero podría estar vinculada al índice de precios al consumo, y en cualquier caso existe una gran sospecha de que con el sistema actual, cuando el precio del mercado sube en relación con los TPDS, se produce un aumento de las pérdidas, con lo que los pobres siempre terminan teniendo menos.

Desde luego quedaría resolver el problema de cómo asegurarse de que el dinero llegue realmente a quienes se ha destinado, pero aquí es donde la informática puede sernos de ayuda. Suráfrica fue el primer país en aplicar la tecnología del uso de cajeros automáticos capaces de reconocer huellas dactilares para pagar las pensiones, y un sistema de este tipo podría funcionar muy bien en la India. Sin duda alguna, merece la pena ponerlo a prueba.

Sin embargo, no está muy claro que un programa de transferencia de efectivo, por muy bien que se lleve a cabo, mitigue de manera efectiva el problema de la malnutrición. Como han señalado recientemente Deaton y Dreze (2008), el incremento considerable de la renta de los pobres que se produjo entre 1983 y 2004 no se tradujo en un aumento del consumo de calorías o proteínas, ni siquiera dentro del grupo que sobrevive con 1.600 calorías diarias. De hecho, hubo una disminución en el consumo de ambas en todos los otros grupos (menos pobres).

Todo ello hace pensar que es posible que los pobres no gasten lo que es debido en alimentación, ya sea porque no la valoran lo suficiente, o porque no quieren perderse por completo el paraíso del consumidor en el que se está convirtiendo la clase media india.⁴ Cualquiera de las dos posibilidades sugiere que una parte importante de las medidas políticas relativas a la alimentación quizá deba consistir en informar e incluir los hábitos de consumo de la población. Hay otra prueba que sustenta esta argumentación: la lactancia materna hasta los seis años es una forma sencilla y muy recomendada de combatir la malnutrición así como muchas enfermedades infantiles. Se recomienda empezarla inmediatamente después de que el niño haya nacido, de manera que éste aproveche el calostro, de alto contenido nutricional. Sin embargo, sólo una cuarta parte de las madres que aparecían en el NFHS declararon haber empezado a dar de mamar al niño una hora después del parto.

La dificultad aquí es modificar la conducta, incluidas formas de comportamiento que se encuentran profundamente arraigadas en las tradiciones. La idea actual del gobierno indio es traspasar esta responsabilidad a una figura llamada ASHA Sahayogini, una mujer con un cierto grado de formación a la que se concederán 23 días para educar los hábitos de las mujeres de su comunidad y una remuneración de 25 dólares estadounidenses al mes. No está aún claro que las mujeres que asuman este trabajo dispongan de la energía, los conocimientos o el carisma necesarios para convencer a otras mujeres de que renun-

⁴ Éstas no son las únicas explicaciones posibles. Deaton y Dreze (2008) sugieren que el bajo consumo en alimentación puede ser un reflejo de que la demanda de trabajo físico es menor.

cien a sus muy arraigadas costumbres. Ahora bien, tampoco parece probable que se vaya a realizar una evaluación seria de este programa en un futuro próximo.

La educación

El bajo rendimiento escolar de la educación primaria de la India ha sido noticia en los últimos años gracias a los informes de las inspecciones anuales de educación publicados por la prestigiosa ONG educativa Pratham. Los puntos básicos de estos informes son bien conocidos: el 42% de los alumnos de quinto grado en la India no tiene el nivel de lectura correspondiente al de segundo grado, y un 68% no sabe hacer restas de dos cifras.

Ahora bien, aunque existen colegios en los que hay más de 100 niños por aula, el sector educativo de la India no está infradotado, si se lo compara con otros países similares. En 2005 la India destinó un 3,7% de su PIB a inversiones en educación, lo cual está un poco por debajo de la media para los países de renta media-baja (4,3%), pero por encima de la media de la región de Asia Oriental-Pacífico (2,9) (Banco Mundial 2007).

Según un trabajo reciente realizado por Murgai y Pritchett (2007), los profesores estatales de la India están mejor pagados que otros profesionales de categoría similar. La proporción media alumno/profesor es elevada, si bien inferior a 40, que es el límite que existe en Israel (un país mucho más rico).

Parece que el problema, al menos en parte, radica en la calidad de la enseñanza. Según los datos del Estudio mundial de absentismo laboral (Chaudhury, Hammer, Kremer, Muralidharan y Rogers 2003), en el que se envió a supervisores a las escuelas a determinadas horas escogidas al azar a fin de que midieran el grado de presencia de los profesores, el 25% de los profesores no está en el colegio diariamente. Además, en caso de que estén allí, sólo pasan en el aula un 45% del tiempo que se supone deben dedicar a dar clase.⁵

Igualmente asombroso, aunque se le concede menos atención, es la tasa de absentismo escolar infantil, comparable o superior a la de los profesores. Dada esta situación cabe preguntarse si los profesores no están simplemente reaccionando a la situación general de indiferencia que cunde entre los alumnos: quizás, después de todo, el que los profesores asistieran unos cuantos días más a clase no tendría un gran efecto en el rendimiento escolar de los alumnos. Un experimento aleatorio reciente que está recogido en Duflo, Hanna y Ryan (2007) ha puesto a prueba esta hipótesis. Seva Mandir, una importante ONG que opera en el estado de Rajastán, tenía unas tasas de absentismo entre los profesores del 40% en escuelas de un solo maestro situadas en zonas remotas. Animados por Duflo, empezaron a controlar la presencia de los profesores por medio de cámaras y les pagaron según el número de días que fueran a trabajar. Este plan se introdujo en una serie de escuelas de manera aleatoria, de modo que se pudiesen evaluar los efectos.

Muchas personas de la comunidad de Seva Mandir consideraron que si bien esto podría hacer que los profesores acudieran más a sus puestos de trabajo ello no repercutiría en el rendimiento escolar. Pero, de hecho, hizo que las puntuaciones de las pruebas de evaluación aumentasen en una variación estándar nada despreciable del 0,17, lo cual puso de manifiesto que los alumnos salían beneficiados cuando los profesores se esforzaban más.

Muralidharan y Sundararaman (2006) también llegaron a la conclusión de que si los profesores recibían mayores incentivos el rendimiento de los alumnos mejoraría. En su estudio incluyeron un experimento realizado en Andhra Pradesh, donde se prometió recompensar a los maestros de escuelas estatales que mejoraran el rendimiento de sus alumnos y comprobaron que los alumnos mejoraban su puntuación en los exámenes, incluso en aquellas materias que no contaban para los incentivos.

No obstante, el efecto de los incentivos fue, de nuevo, no muy grande, con una variación estándar del 0,15. Sin duda alguna habrá que tomar muchas más medidas para transformar el hoy por hoy deficiente sector educativo de la India. Pero ¿cómo generamos los incentivos necesarios para que esto ocurra?

Una respuesta, que constituyó una de las cuestiones centrales del último mayor intento realizado por el gobierno de la India para reformar la educación primaria, la Sarva Shiksha Aviyan (o SSA), es que la comunidad debe desempeñar un papel mucho más activo a la hora de exigir educación. Ahora bien, en el estudio realizado en 280 pueblos en el distrito de Jaunpur del estado de Uttar Pradesh se puso de manifiesto que por lo menos cuatro años después del inicio de la SSA, el 92% de los padres no parecían estar enterados de la existencia de los llamados Comités de educación de los pueblos (VEC) —su intermediario principal para participar en la mejora de las escuelas locales y tener acceso a los fondos de la SSA— y sólo un 2% eran capaces de nombrar a sus miembros (Banerjee, Banerji, Duflo, Glennerster y Khemani 2006). Entonces supusimos que ello se debía a que nadie se había tomado la molestia de informarles sobre la VEC o la SSA. Por consiguiente llevamos a cabo un experimento de campo en el distrito encaminado a informar y a movilizar a los padres sobre la educación en sus pueblos y las nuevas posibilidades derivadas de la SSA.⁶ En este experimento, voluntarios de Pratham pasaron un día y medio en cada aldea y mantuvieron pequeñas y grandes reuniones en las que informaban a los padres sobre sus derechos, incluidos el derecho de quejarse de los profesores que no asistieran a clase y el de contratar asistentes adicionales para los profesores o *shikshakarmis* (para las escuelas en las que hubiese demasiados alumnos). Asimismo, les hablaron del (bajo) rendimiento escolar de sus hijos y les enseñaron a evaluar su capacidad lectora.

Nada de esto tuvo ningún efecto en los padres aparte del hecho de que ahora un 2,6% sabía que existían los

⁵ Ello coloca a la India entre los países que peor parados salen del estudio (sólo Uganda, con un 27%, tiene un índice mayor de absentismo).

⁶ Para el experimento véase Banerjee, Banerji, Duflo, Glennerster and Khemani (2008).

VEC, lo cual supone una diferencia significativa estadísticamente, pero en cualquier caso mínima. No hubo más quejas, ni más visitas a las escuelas, ni se intentó contratar a más asistentes para los profesores. Y, lógicamente, las puntuaciones de los alumnos en las pruebas de evaluación siguieron siendo las mismas.

Lo que no sabemos es si esta indiferencia tiene su origen en la creencia de que la educación no es el camino a la prosperidad para las personas como ellos (después de todo en la India existe una tradición que afirma que sólo ciertas castas superiores tienen derecho a la educación). O puede que crean que los profesores están fuera de su alcance, desde un punto de vista político y social, con lo que intentar influir en su manera de actuar no constituye una alternativa real. Hay datos recientes que apuntan a una combinación de ambas causas: Jensen realizó un experimento en República Dominicana en el que explicó a una serie de padres pobres las ventajas de recibir una educación, y comprobó que, a consecuencia de ello, sus hijos trabajaban mejor en el colegio (Jensen 2007).

Por otro lado, la única intervención satisfactoria de nuestro estudio del estado de Uttar Pradesh se produjo cuando los instructores de Pratham enseñaron a voluntarios de los pueblos a dar clase. Se pusieron en marcha una o más clases en cada uno de los pueblos, y los niños que asistieron a ellas mejoraron de manera considerable su rendimiento en las pruebas de evaluación. El éxito de este experimento así como el fracaso de las otras intervenciones (todas en las que era necesario un cierto grado de interacción social) indican que a los padres sí les importa la educación de sus hijos, pero les da miedo enfrentarse a los profesores.

En ambos casos es difícil tener la certeza de que un mayor grado de participación por parte de los padres vaya a resolver la falta de incentivos, al menos en un futuro próximo. La alternativa son las soluciones de mercado, es decir, algún tipo de programa en el que se dé a los padres cupones financiados por el Estado para pagar las matrículas de los colegios privados de sus hijos. El argumento habitual en contra de estos cupones no parece muy convincente en el contexto indio: ¿generará más segregación y desigualdad en la educación que reciben los niños? Es posible, pero dado que la élite rural ya ha abandonado el sistema estatal en muchas áreas, como mínimo es verosímil que ayude a paliar la desigualdad, al menos siempre y cuando estos cupones se diseñen de tal manera que no puedan usarse para enviar a los niños a colegios realmente elitistas. ¿Debe preocuparnos la posibilidad de que los padres pacten en secreto con la dirección del colegio el cobro de los cupones en efectivo en lugar de enviar a sus hijos a clase? Es poco probable, en nuestra opinión, ahora que a los padres les importa lo suficiente la educación como para que se alcancen unas tasas de participación escolar de casi el 100%.

Además, el verdadero auge de la educación privada entre las familias relativamente pobres de las áreas rurales de la India en los últimos años se traduce en que en

muchos pueblos hay muchos colegios privados que compiten por los alumnos. Según ASER (2007), el 19,3% de todos los niños entre 6 y 14 años de estas regiones estudia en colegios privados. Muralidharan (2006) informa sobre un estudio representativo a escala nacional relacionado con las escuelas primarias privadas de las zonas rurales y observa que un 50% de los colegios identificados en el estudio de 2003 se habían fundado en los últimos cinco años.

Muralidharan también señala que estas escuelas son baratas (el precio medio mensual es de 2 dólares estadounidenses al cambio actual) a pesar de que es más probable que los profesores que trabajan en ellas sean licenciados universitarios y que su proporción profesor/alumno es casi la mitad de la de los colegios públicos. Esto se debe a que los profesores de los colegios privados cobran entre una décima y una quinta parte de lo que cobran los de los colegios públicos. Andrabi et al. (Andrabi, Khwaja y Das 2003) examinaron un fenómeno similar en la provincia de Punjab en Paquistán y sostienen que la diferencia de los resultados entre colegios privados y públicos es demasiado grande como para que se pueda explicar con una serie de razonamientos convincentes o, lo que es lo mismo, que las escuelas privadas son sencillamente más baratas y mejores.

No obstante, es evidente que necesitamos pruebas más fehacientes antes de justificar una diferencia tan radical. Karthik Muralidharan y Michael Kremer están realizando en este momento una evaluación aleatoria de los cupones escolares en el estado de Andhra Pradesh; esperemos que ocurra lo mismo con otros programas de cupones que se pongan en marcha en otros estados en un futuro. La dificultad que tienen estas evaluaciones reside en hacer frente al hecho de que la provisión de colegios privados tendrá que ajustarse a la expansión de la demanda que se producirá una vez que el sistema de cupones se generalice, pero esto no ocurre en condiciones experimentales. Lo que se teme es que las tarifas subirán de manera drástica, pues a los colegios les resultará muy difícil encontrar a personas que den clase. A fin de dar respuesta a estas preguntas, Muralidharan y Kremer hicieron selecciones aleatorias por los pueblos y dentro de ellos. Si el pueblo ofrece buenas oportunidades laborales para los profesores, entonces el experimento realizado nos dará estimaciones sobre precios y número de colegios. Si por el contrario los profesores están dispuestos a mudarse de pueblo a fin de encontrar un empleo, lo cual parece probable, los datos obtenidos no serán completos y serán necesarios nuevos estudios.⁷ Entretanto, el sector de la educación va claramente a la deriva.

Atención médica

La atención médica adolece de los mismos problemas que la educación, pero en mayor escala. La tasa de absentismo laboral entre las comadronas auxiliares (ANM), que constituyen la categoría más baja de profesionales médicos del jerárquico sistema sanitario de la India, es del 40%, bas-

7 La respuesta también puede que dependa de que la ampliación del programa de cupones coincida con el cierre de un gran número de colegios, en cuyo caso habrá muchos profesores disponibles.

tante más elevada que la de los profesores (Chaudhury, Hammer, Kremer, Muralidharan y Rogers 2003). Cuando se escogieron al azar una serie de centros secundarios de salud (donde hay enfermeras) para un programa de incentivos basados en la asistencia, éstas y sus jefes inmediatos pactaron para inutilizar por completo los incentivos: la asistencia de las enfermeras tras el experimento fue tan baja como lo era antes del mismo (por debajo del 40%) (Banerjee, Duflo y Glennerster 2008).

Aún más preocupante, aunque quizás no tan sorprendente dadas las tasas de absentismo laboral, es el hecho de que hasta las personas muy pobres hayan dejado casi por completo de acudir a estas enfermeras. En el distrito rural de Udaipur, donde el gasto medio diario per cápita no supera un dólar, encontramos que menos de una cuarta parte de las visitas a profesionales que ofrecen servicios médicos tenían lugar en hospitales o ambulatorios del Estado. Casi un 60% de las visitas eran a profesionales privados y el resto a curanderos (Banerjee, Deaton y Duflo 2004). Esto ocurre a pesar de que los médicos «privados» están más lejos, son más caros y es probable que tengan menos acreditaciones.

Cuando preguntamos a posibles pacientes por las razones de este fenómeno, señalaron que se debía a la calidad del tratamiento. Sabemos que la calidad es con frecuencia deficiente, pues ya hemos hablado de las elevadas tasas de absentismo. Asimismo, Das y Hammer (2007), basándose en un estudio sobre médicos estatales y privados en las zonas urbanas de Delhi, señalan que los médicos estatales que se ocupan de los pacientes más pobres a menudo prescriben tratamientos sin examinarlos siquiera. Sin embargo, una parte de lo que los pacientes denominan calidad es precisamente de lo que se quejan los profesionales que trabajan para el Estado, que afirman que los médicos privados abusan de los medicamentos inyectables, en especial de los antibióticos y los esteroides, algo que la población parece considerar un buen tratamiento. Nuestros datos apoyan en cierto modo esta visión. Un 60% de todas las visitas a profesionales privados lleva aparejada una inyección, ahora bien, no disponemos de datos que indiquen si éstas son realmente peligrosas para los pacientes. El consenso general entre los expertos es que la sobremedicación es considerable.

El desplazamiento hacia la atención médica privada trae un problema añadido: los pacientes ya no reciben información por parte del Estado sobre prevención y prácticas saludables. Por ejemplo, en el distrito rural de Udaipur menos del 5% de los niños están convenientemente vacunados según nuestros datos (Banerjee, Duflo, Glennerster y Kothari 2008), y no parece que los profesionales o centros que ofrecen atención médica privada estén dispuestos a hacer algo al respecto.

Más en general, lo que hace que la asistencia sanitaria sea aún más difícil para los pobres es que las soluciones que proporciona el mercado no son necesariamente

muy atractivas, sobre todo por esta tendencia a despreciar los económicos si bien valiosos aspectos preventivos de la medicina frente a tratamientos caros y quizás perjudiciales. Los seguros médicos subvencionados son el equivalente de los cupones para la educación, y hay una serie de experimentos en curso en la India, incluido uno que nosotros estamos evaluando. No obstante, casi todas estas pólizas de seguros sólo cubren los gastos de pacientes hospitalizados, por la sencilla razón de que se pueden comprobar con mayor facilidad. Todo ello supone que los chequeos médicos, las pruebas y otras prácticas preventivas continúen siendo un gasto que las personas tienen que pagar de su propio bolsillo, y que los seguros a menudo desaconsejan estas prácticas (porque adoptarlas reduce los costes en caso de enfermedad).

En el momento actual existen dudas sobre si la atención médica subvencionada es viable económicamente. Es probable que un programa que cubra más pacientes externos sea mucho más caro, debido a que el uso de estos servicios es más difícil de supervisar y puede hacer necesaria la intervención del gobierno. También se podría utilizar el programa a modo de cebo, para hacer que más personas participen a la hora de detectar y prevenir con tiempo las enfermedades (esto es, que para conseguir seguros médicos subvencionados tengan que cumplir una serie de requisitos).

El hecho de que estos incentivos pueden funcionar ha quedado demostrado por un estudio experimental reciente, en el que se ofrecía un kilo de lentejas a las mujeres cada vez que llevasen a sus hijos a vacunar y que multiplicó por encima de dos el porcentaje de niños totalmente vacunados (Banerjee, Duflo, Glennerster y Kothari 2008).

Las medidas adoptadas por el gobierno en la India se están encaminando en este sentido. En la actualidad existe un plan que ofrece incentivos económicos a las mujeres que den a luz en el hospital, y, como parte de este plan, también tienen que visitar la clínica un número determinado de veces antes y después del parto. Aunque el cumplimiento de estas nuevas reglas parece de momento poco estricto, podría llegar a ser determinante.

El camino a seguir

La tendencia actual de las medidas adoptadas para combatir la pobreza se aleja bastante de la idea de que los pobres tienen que hacerse cargo de sus vidas. En lugar de entender que son trabajadores que necesitan adquirir una serie de destrezas, se les considera empresarios potenciales que necesitan derechos sobre el capital y la propiedad así como el amparo de la ley. De ahí que se dé importancia al microcrédito, por ejemplo. Esto no quiere decir que crean que no es importante invertir en capital humano; más bien los defensores de esta perspectiva se muestran bastante escépticos con respecto a la capacidad del gobierno a la hora de aportarlo, por lo que prefieren que

los pobres reciban ingresos adicionales y los utilicen para pagar el capital humano que necesitan para sus hijos.

El hecho de que no es fácil conseguir que el gobierno se encargue de esto concuerda con el razonamiento que estamos presentando. La cuestión es si podemos estar seguros de que en los casos en los que el gobierno no lo haga, los pobres puedan conseguirlo por sí solos, es decir, que sean capaces de salir adelante sin recurrir a nadie, únicamente con una pequeña ayuda procedente de las organizaciones de microcrédito.

Como hemos expuesto ampliamente en otros trabajos (véase Banerjee y Duflo 2007, y Banerjee y Duflo 2008) no existe una justificación empírica de esta perspectiva. Lo fundamental es que los pobres no tienen ni las aptitudes ni el conocimiento de los mercados, ni la comprensión ni la tecnología necesarias para competir de manera eficaz en este entorno. Incluso tras haber obtenido un microcrédito, su falta de medios económicos les da acceso únicamente a tecnologías rudimentarias y les destina a ejercer las profesiones más multitudinarias. Además, son enormemente vulnerables a todos los riesgos que pueda correr cualquier otro empresario. Todas estas limitaciones son evidentes en los negocios que emprenden: son muy pequeños (la empresa media de un propietario pobre no tiene empleados) y están muy concentrados en unos pocos sectores en los que es posible trabajar sin destrezas especiales ni especialización.

Lo que es más, los propios pobres no esperan que sus negocios transformen sus vidas. Si lo hicieran se esforzarían más en hacer que éstos crecieran más de lo que lo hacen. En nuestra opinión muchos de ellos podrían expandir sus negocios con facilidad y ganar algo más de dinero, con lo que podrían salir de la pobreza, pero eligen no hacerlo.

Con ello no queremos decir que los pobres no sean personas con recursos y energía, sino simplemente que las normas del juego les son tan adversas que sólo los que estén completamente decididos y/o tengan un gran talento consiguen ir más allá de la línea de salida. Tampoco cabe duda alguna de que el microcrédito ha hecho que sus vidas sean menos insoportables, con lo que merece nuestro apoyo. Ahora bien, en última instancia es el gobierno el que debe permanecer en el centro de las medidas para combatir la pobreza, pues sin ayuda ni recursos externos los pobres tienen que enfrentarse a un desafío en condiciones injustas. No es necesario que el Estado continúe haciendo todo lo que hace (mal) en la actualidad, y, sin duda debería centrarse más en pagar servicios en lugar de proporcionarlos: los subsidios complementarios pueden hacer mucho para mejorar la vida de los pobres sin que el gobierno tenga que participar en el proceso. Con todo, no debemos olvidar que una parte muy importante de lo que éste hace se refiere a aspectos de los que los mercados no se ocuparían (los cambios de comportamiento, la asistencia sanitaria preventiva, la educación para quienes viven en áreas donde no hay colegios privados, la ayuda en las situaciones de emergencia, etcétera). Incluso en estos casos el gobierno puede trabajar con socios encargados de la ejecución de proyectos externos, como ha demostrado el caso del BRAC, en Bangladesh, pero, si somos realistas, continuará siendo uno de los agentes sociales principales en la economía. El reto que tenemos algunos de los que nos dedicamos a lo que se podría denominar el sector de la ideas, es, por consiguiente, hallar la manera de rediseñar las fórmulas de actuación de los gobiernos para conseguir que funcionen mejor en términos de ejecución y de delegación a terceros, así como de mejorar la eficacia de sus actuaciones futuras.

Bibliografía

- Andrabi, T., A. Khwaja y J. Das. «Learning and Achievement in Pakistani Education». Sector Feasibility Report, Banco Mundial, 2003.
- Annual Status of Education Report (ASER) (2007), «Rural», Centro de Recursos de Pratham, <http://www.pratham.org/aser07/ASER07.pdf>.
- Atanassova, A., M. Bertrand y S. Mullainathan. «Misclassification in targeted Programs: A Study of the Targeted Public Distribution System in Karnataka, India». Universidad de Harvard (inédito).
- Banco Mundial. «World Development Indicators, 2007: Table 2.9, Education Inputs». 2007. Disponible en: http://siteresources.worldbank.org/DATASTATISTICS/Resources/table2_9.pdf.
- Banerjee, A. *Making Aid Work*. Cambridge, Massachusetts: MIT Press, 2007.
- Banerjee, A. y E. Duflo. «The Economic Lives of the Poor». *Journal of Economic Perspectives* 21 (1) 2007: 141-167.
- «What is Middle Class about the Middle Classes Around the World?». *Journal of Economic Perspectives* 22 (2) 2008: 3-28.
- A. Deaton y E. Duflo. «Health Care Delivery in Rural Rajasthan». *Economic and Political Weekly* 39 (9) 2004: 944-949.
- E. Duflo y R. Glennerster. «Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System». *European Economic Association* 6 (2-3) 2008: 487-500.
- E. Duflo, R. Glennerster y D. Kothri. «Improving Immunization Coverage in Rural India: A Clustered Randomized Controlled Evaluation of Immunization Campaigns with and without Incentives». Massachusetts Institute of Technology, 2008 (inédito).
- R. Chattopadhyay, E. Duflo y J. Shapiro. «Targeting Efficiency: How well can we identify the poor?». MIT (inédito).
- R. Banerji, E. Duflo, R. Glennerster y S. Khe-mani. «Can Information Campaigns Spark Local Participation and Improve Outcomes? A Study of Primary Education in Uttar Pradesh, India». *World Bank Policy Research Working Paper*, 3.967, 2006.
- «Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India». *The World Bank Policy Research Working Paper*, 4.584. Impact Evaluation Series No. 21, 2008.
- Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan y H. Rogers. «Teachers and Health care providers Absenteeism: A multi-country study, Mimeo, Development Research Group». Banco Mundial (inédito), 2003.
- Deaton, A. y J. Dreze. «Nutrition in India: facts and interpretations». Universidad de Princeton (inédito), 2008.
- Dreze, J., R. Khera y Siddhartha. «Corruption in NREGA: Myths and Reality». *The Hindu*, 22 de enero de 2008. Disponible en: <http://www.hindu.com/2008/01/22/stories/2008012254901000.htm>.
- Duflo, E., R. Hanna y S. Ryan. «Monitoring Works: Getting Teachers to Come to Work». *Poverty Action Lab*, 2007.
- Filmer, D. y L. Pritchett. «Estimating Wealth Effects without Expenditure Data--Or Tears: An Application to Enrollments in States of India». *Demography* 38.1 2001: 115-132.
- Jensen, R. «The Perceived Return to Education and the Demand for Schooling». Universidad de Brown University (inédito), 2007.
- Leonard, K., J. Hammer y J. Das. «The quality of medical advice in low-income countries». *Policy Research Working Paper Series* 4.501, Banco Mundial, 2008.
- Muralidharan, K. «Public-Private Partnerships for Quality Education in India». (borrador inédito), 2008.
- y V. Sundararaman. «Teacher Incentives in Developing Countries: Experimental Evidence from India». *Job Market Paper*. Universidad de Harvard, 2006.
- Pritchett, L. y R. Murgai. «Teacher Compensation: Can Decentralization to Local Bodies Take India From Perfect Storm Through Troubled Waters to Clear Sailing?». *India Policy Forum* 2006/07, 2007.
- Strauss, J. y D. Thomas. «Health, Nutrition and Economic Development». (En colaboración con Duncan Thomas), *Journal of Economic Literature*, 36, 2 1998: 766-817.

la averiguación artística del futuro

FRANCISCO CALVO SERRALLER

Habiendo surgido de una revolución hace aproximadamente dos siglos y medio, es muy difícil pronosticar cómo será el futuro del arte contemporáneo cuando aún nos hallamos averiguando qué es. Lo que tradicionalmente se entendía como arte hasta llegar a la segunda mitad del siglo XVIII revalidaba lo que al respecto afirmaron los griegos, sus inventores. Hacia el siglo VI antes de Cristo, los griegos definieron el arte como una imitación selectiva de la realidad o la naturaleza; esto es, no una simple copia o réplica indiscriminadas de lo que veían a su alrededor, sino sólo la que, trascendiendo las apariencias, captase el orden subyacente a ellas. La plasmación de este orden producía belleza, con lo que estaba claro que el fundamento y el fin del arte era, pues, la belleza. Pero no una belleza interpretada de forma subjetiva, sino codificable objetivamente, de naturaleza matemática. De esta manera las cualidades estéticas exigibles a una obra de arte clásica respondían siempre a conceptos matemáticos como «armonía», «proporciones», «simetría», «ritmo», etc., todos pautas de un orden. Pero esta selección formal ordenada también tenía su correspondencia en relación con el contenido o tema representados, porque tampoco estaba autorizado que éstos se pudiesen escoger al arbitrio de su autor, sino que debían ser socialmente edificantes; esto es, debían ajustarse a la acción aleccionadora protagonizada por dioses y héroes de un remoto pasado

mítico. En suma, los griegos definieron el arte como, por así decirlo, una buena representación del bien, donde se conjugaban el orden formal y el orden moral. Pues bien, este canon artístico estuvo vigente, en el seno de la civilización occidental, durante aproximadamente veinticuatro siglos, creando una tradición histórica conocida unitariamente como «clasicismo». Es evidente que esta tan prolongada tradición padeció sucesivas crisis, cuya gravedad, a veces, abrió algunos amplios paréntesis como el muy significativamente denominado «Edad Media», donde, sin desaparecer por completo, los principios artísticos clásicos estuvieron muy corrompidos, pero precisamente por ello, al comienzo de la época moderna se utilizó la fórmula restauradora del Renacimiento, que lo era de los valores y criterios de la Antigüedad clásica grecorromana. A partir del siglo XV, esta restauración del ideal artístico clásico, tomada al principio con dogmático entusiasmo, sufrió no pocos embates polémicos al hilo del discurrir de los profundos cambios de toda índole que deparó el mundo moderno, pero consiguió resistir en su esencia doctrinal justo hasta los albores de nuestra revolucionaria época, que cambió de forma radical lo que hasta entonces se entendía como arte.

Tras esta sucinta, sintética y reductora versión de la historia artística del clasicismo occidental, se puede entender quizá mejor la revolución subsiguiente. El primer cariz de

esta revolución artística fue insinuar que el fundamento exclusivo del arte no sólo no podía ser la belleza, sino que ésta constituía una barrera insoportable para el potencial desarrollo de aquél, que exigía un horizonte mucho más amplio. Fue entonces, a lo largo del siglo XVIII, cuando se buscaron ideales para el arte distintos y contrapuestos a la limitadora belleza como lo «sublime» o lo «pintoresco», que postulaban que el «desorden» y la «irregularidad» podían proporcionar al hombre experiencias estéticas nuevas, más íntimas y reveladoras que las tradicionales.

En el primer párrafo del capítulo tercero de su obra *Laocoonte o sobre las fronteras de la poesía y la pintura*, publicada en Berlín en 1776, su autor, G. Ephraim Lessing, escribió lo siguiente: «Pero, como ya hemos observado, en los últimos tiempos el arte ha adquirido dominios incomparablemente más vastos. El campo donde se ejerce su imitación se ha extendido a la Naturaleza entera visible, y de ésta lo bello es solamente una pequeña parte. La verdad y la expresión, se dice, son la expresión suprema del arte; y del mismo modo como la Naturaleza está sacrificando continuamente la belleza en aras de designios más altos, asimismo el artista debe subordinar esta belleza a su plan general, sin buscarla más allá de lo que le permiten la verdad y la expresión. En una palabra, la verdad y la expresión transforman la fealdad natural en belleza artística». Entre otros muchos testimonios contemporáneos, la clara y rotunda explicación que nos proporciona este texto de Lessing nos enfrenta de una forma inapelable con el problema del revolucionario arte de nuestra época, que ya no será un arte basado en la belleza. Más aún, se puede afirmar que las polémicas estéticas y el desarrollo del arte mismo durante la segunda mitad del siglo XVIII se rigen por el común afán de librar una auténtica guerra de liberación contra la belleza, cuyo canon clásico es asaltado con el fervor que se empleó en la toma de la Bastilla.

Ahora bien, ¿cómo fundamentar de nuevo el arte despojado de su tradicional cimentación? Ni el placer por lo ilimitado de lo sublime, ni por lo irregular de lo pintoresco son en sí mismos fundamentos, sino desafíos al concepto tradicional de belleza. En cualquier caso es evidente que cambian la perspectiva, un poco al modo de giro copernicano en la física clásica, porque recomponen o resitúan el horizonte que ya no se centra en el consabido más acá, sino en un desconocido más allá. ¿Cómo abarcar ese cosmos estético desconocido sin imponerle unos restrictivos límites como los de la belleza? Por otra parte, es importante retener que esta rebelión contra el dominio hegemónico de la belleza clásica nunca significó que los partidarios de la revolución artística anulasen el valor del arte tradicional, sino que se obligase a su cumplimiento universal; esto es, que se negase valor artístico a los que querían explorar nuevas sendas. En suma, al tomarse conciencia de que el horizonte artístico podía ser indefinidamente más amplio que el clásico, pero todavía sin haber encontrado unos límites precisos que acotasen su nueva

dimensión, es lógico que la cuestión se zanjase de la forma más abierta y negativa posible, es decir, buscando el fundamento de lo artístico en lo infundamentado, en la libertad. Es lo que formula Schiller, ya en la crucial última década del siglo XVIII cuando, en sus *Cartas sobre la educación estética del hombre* afirma que es propio del estadio estético «dar la libertad por medio de la libertad», o, lo que es lo mismo, que el fundamento del nuevo arte revolucionario sería, en efecto, la libertad, cuya única circunstancial limitación no es estética sino «técnica».

Si el único fundamento que cabe poner al arte es que a través de él se accede a lo libre que por fuerza niega cualquier limitación anterior, ¿cómo cartografiar el nuevo mapa del arte sino de la forma obvia que señala que su territorio es el que van descubriendo sus sucesivos exploradores, o, lo que es lo mismo, que «arte es lo que llamamos arte» porque así nos lo indican los artistas y nosotros lo damos por bueno? Quien, al margen de estas apresuradas disquisiciones generalistas, observe la evolución material del arte desde la segunda mitad del siglo XVIII hasta ahora mismo, comprobará que no es que el arte haya ido cambiando continuamente de forma o estilo, sino que, cada vez, al hacerlo, ha redefinido de nuevo lo que es arte. Significativamente, si analizamos las reacciones del público artístico ante cada nuevo cambio experimentado por el arte en nuestra época, veremos, la mayor parte de las veces, que no niegan el valor de su forma o de su contenido, sino que eso sea arte. La postura contraria no modifica tampoco nada porque cuando, como ahora, el público acepta todo lo nuevo porque es nuevo, indefectible y simétricamente está reconociendo que, de entrada, todo puede ser arte, es decir, que si nada es arte o todo es arte, es el arte en sí el que constantemente está puesto en cuestión.

Hasta cierto punto, es verdad que no fue igual la perplejidad producida por el arte nuevo cuando éste todavía utilizaba soportes y técnicas tradicionales, pero no podemos atribuir a ello, como a veces se hace, el quid de la cuestión. Quiero decir que ni la postulación de que un urinario de fabricación industrial firmado por un artista, ni el desarrollo de nuevas técnicas como la fotografía o el cine fueran rechazadas o aceptadas resuelven nuestra perplejidad, que es producto de tener que descubrir constantemente qué es arte. Esta incertidumbre ha provocado una polarización social entre quienes descalifican o defienden absolutamente el arte contemporáneo, los primeros agarrándose al concepto tradicional del arte como su única manifestación genuina, o, lo que es lo mismo, no mirando más que al pasado, mientras los segundos simétricamente postulando que sólo un arte innovador es el que verdaderamente responde a las exigencias del hombre actual que lo cifra todo en la expectativa del futuro. Ahora bien, ¿es indiscutiblemente cierto que se haya abierto un abismo insalvable entre el arte tradicional y el de nuestra época? ¿Es sólo el del pasado el auténtico arte y el contemporáneo una simple superchería o viceversa? La ansiedad lleva

al hombre a afirmaciones tajantes, cuyas formulaciones teóricas diáfanas no lo son tanto cuando se confrontan con la realidad vivida. Y aunque, desde luego, siempre es más fácil explicarse el pasado que el presente, que el arte contemporáneo no responda al canon clásico no significa que su práctica sea arbitraria, sino, en todo caso, no reglamentada de manera objetiva. Basado en la anticánónica y elástica libertad, el arte de nuestra época está necesariamente volcado a la exploración, que no se detendrá hasta que encuentre un límite, el cual, a su vez, permitirá fijar las nuevas fronteras, pero es evidente que no hemos llegado a este momento. De todas formas, los dos siglos y medio de práctica artística contemporánea han dejado un curso histórico susceptible de ser comprendido y ordenado, lo que indica que la incertidumbre se sitúa siempre en el presente pero no en el pasado, que es también un pasado moderno, o, como se suele decir al respecto, que hay ya una sólida tradición vanguardista. Las sucesivas novedades o modas artísticas han pasado un filtro crítico que, como en la anterior etapa clasicista, ha dejado fuera de la memoria gran parte de lo realizado, es decir, que sigue operativa la selección de lo mejor, que sólo se revalida con el paso del tiempo. De manera que, el hecho de que no sepamos adónde va el arte no ha supuesto el imperio de la indiscriminación, entre otras cosas porque salirse de un camino trillado no lleva necesariamente a perderse, porque entonces no se habría producido jamás ningún descubrimiento en nada. Para terminar con esta cuestión hay además una prueba histórica irrefutable: que ninguno de los artistas contemporáneos más revolucionarios, desde Goya a Picasso, han dejado de innovar sin tener en cuenta el pasado. Y aunque de una forma sucesivamente diferente, esto sigue siendo hoy así, como se pone de manifiesto a través de incontables ejemplos.

Por otra parte, ¿no está ocurriendo lo mismo en todos los campos contemporáneos del conocimiento y de la experiencia? Es cierto que el hoy hegemónico conocimiento científico y, aún mejor, su excrecencia técnica, se rigen por una dinámica de progreso, cada uno de cuyos nuevos pasos invalida los anteriores, pero no está claro que esta ideología se pueda aplicar al terreno del arte, cada uno de cuyos ámbitos e incluso los de índole revolucionaria nunca han implicado la invalidación de lo anterior. En realidad, el aprecio de Duchamp no ha restado valor a Fidias o a Rafael, ni el de éstos debería descalificar a quienes no siguiesen repitiendo su fórmula. El arte, así pues, hasta el presente, ha cambiado de manera constante, pero sin que esa acumulación de novedades anule el valor a las antigüedades. ¿No habrá que considerar, por tanto, esa radical contraposición entre el arte de hoy y el de ayer una falacia desmentida constantemente por los hechos? Por lo demás, que no tengamos ninguna garantía de que ocurra lo mismo en un indeterminado e indeterminable futuro tampoco nos autoriza a adoptar una perspectiva apocalíptica que, de producirse, será, en todo caso, más grave

en cualquier otra dimensión que en la artística, cuya área de acción es comparativamente inocua. En realidad, por lo general, el arte propende más que cualquier otra práctica a establecer puentes históricos, no sólo porque no responde a un modo de conocimiento unívoco sino porque es una constante reflexión indagatoria sobre la experiencia, lo vivido, lo pasado, y su móvil existencial principal es hacerse duradero, memorable, inmortal.

Pero ¿cuál ha sido la huella histórica que ha dejado el arte de nuestra época durante sus aproximadamente dos siglos y medio de vigencia? Es difícil discernirla si no se entiende previamente el motor modernizador que la ha regido. El término «moderno» etimológicamente procede del latín y significa lo «actual», con lo que modernizar el arte supone «actualizarlo»; esto es, hacerlo girar en torno al presente tal y como se nos presenta, con sus modos o modas, con, en fin, sus novedades o innovaciones. Modernizar el arte clásico o clasicista significó, en primer lugar, modernizar su contenido, su concepto de historia, lo que hasta entonces narraba, que no era el presente y menos lo presente tomado de forma indiscriminada. Según la jerarquía estética establecida por los griegos, el punto más elevado y profundo de una historia era su modo trágico: la celebración renovada de un sacrificio ancestral y fundante cuya rememoración tenía un efecto aleccionador traumático —catártico— sobre el espectador. En el punto más bajo, por su parte, estaba el modo cómico, cuya característica era tratar de las historias contemporáneas de los hombres mortales y su efecto aleccionador era mucho menos intenso y, hasta se puede decir, que «relajante». Trasladado este esquema al mundo de las artes plásticas, se entiende que el género superior fuese el de la narración trágica, mientras que los restantes, subordinados e inferiores, representaban episodios cómicos hasta casi bordear lo insignificante. Teniendo en cuenta lo antedicho, modernizar artísticamente la representación de lo histórico supuso no sólo ocuparse preferentemente de las historias menores de los mortales, ellos mismos carentes de importancia social y significación simbólica trascendente, sino incluso hasta obviar su presencia directa, haciendo que un paisaje o cualquier objeto alcanzasen por sí solos la misma dignidad que la acción humana. En este sentido, la modernización artística del contenido fue similar a su «democratización», haciendo que los temas tratados estuviesen al alcance de cualquiera y su mensaje fuera inmediato, puntual, como cogido al vuelo de lo que pasa sin más. Quien observe este proceso modernizador se percatará de cómo progresivamente es un continuo acoger no sólo las historias comunes, sino lo más común o intrascendente de cada una de ellas hasta, ya en el ecuador del siglo XIX, iniciar la senda de la insignificancia.

Llegado el arte de vanguardia al punto de contar todo, casi el siguiente paso renovador obligado era no contar nada, es decir, «desnarrativizar» o «desliteratular» el arte, cuya identidad auténtica ahora se cifraba en ser For-

ma, o, si se quiere, en todos los elementos materiales que constituyen una obra de arte al margen de lo que represente o simbolice. Esta segunda gesta modernizadora del arte, que concluyó aproximadamente con la Primera Guerra Mundial, fue la que centró su esfuerzo revolucionario en los hasta entonces considerados elementos formales, que constituían lo que el arte era por sí mismo, y su objetivo fue dar al traste, en primer término, con el sistema de representación en perspectiva heredado, y, en segundo, por así decirlo, abordar la «nada» resultante. Porque ¿qué quedaba de una obra plástica sin su contenido y su forma tradicionales? El instrumento quirúrgico para esta operación revolucionaria fue el cubismo, que abocó a un arte anicónico o abstracto, pero además, y casi de inmediato, replanteó por completo lo que se entendía como arte, que ya no requería ser un cuadrado o una escultura, sino cualquier cosa y a través de cualquier medio, como de alguna manera dio a entender ese heteróclito y nihilista movimiento que se denominó dadaísmo. Si tenemos en cuenta que, durante esa procelosa segunda década del siglo xx, además de los *collages* y de los *ready made* también cobraron una arrolladora presencia la fotografía y el cine, se puede comprender que ya se pudiera dar por liquidado por completo, fondo y forma, el arte tradicional. Si la modernización del contenido dejaba abierta la posibilidad para que el nuevo arte careciese de cualquier restricción temática, y si la modernización de la forma permitía en principio que cualquier cosa fuera arte, tal y como ha venido ocurriendo desde entonces hasta hoy, ¿qué límites habría de tener el insondable horizonte artístico resultante, esa «nada musical», como le gustaba comentar a Mallarmé?

Pero ¿qué pasó desde ese momento en que la vanguardia ultimó los postreros residuos del arte tradicional hasta ahora mismo, unos ochenta años después? En cierta manera, se puede afirmar que, tras el salto cualitativo dado por el arte de vanguardia durante el primer cuarto del siglo xx, no se ha producido ninguna innovación esencial más, lo que no significa que no hayan dejado de ocurrir incidencias en el mundo artístico desde el punto de vista social, económico y tecnológico. Tampoco se puede desdeñar el fenómeno estético de la propia autodestrucción de la vanguardia como tal, que no logró sobrevivir durante el último cuarto del siglo xx, que no en balde no cesa de autoproclamarse, de una manera harto equívoca, como un periodo artístico «posmoderno», algo que hay que interpretar como el periodo del definitivo triunfo social de la modernidad y de la inanidad de seguir usando una estrategia de pugna lineal para su afirmación. Hoy, en efecto, el consumo de novedades no necesita de ninguna falange organizada de apoyo. Así parece demostrarlo el que el mercado de arte actual se haya convertido en hegemónico y el que las plataformas institucionales de legitimación, como los museos de arte contemporáneo, no sólo se multipliquen de forma exponencial sino que ellos mismos se hayan transformado en obras artísticas.

Este asunto, ya que sale al paso, merece una cierta consideración monográfica. Antes que nada, hay que recordar al respecto que los museos públicos fueron una creación de nuestra época justo en la transición del siglo xviii al xix, pero que los denominados «de arte contemporáneo» surgieron a finales del xix y se han desarrollado durante el xx. La razón de ser de estos últimos inicialmente fue la respuesta política ante el rechazo social de que el arte sucesivamente nuevo se integrase en los museos históricos, porque producía una auténtica perplejidad pública que un Manet o un Monet se alineasen junto a las consagradas obras de los antiguos maestros. Con la intención de resguardar al nuevo arte del rechazo social se creó el Luxemburgo parisino, que pronto sirvió de pauta para otros países afectados por el mismo problema, como España que, en 1895 creaba su primer museo de arte contemporáneo. Fue entonces cuando se creó una tripartición de los museos públicos, que se dividieron en arqueológicos, históricos y contemporáneos debiendo su identidad los primeros y los terceros significativamente a ocuparse, por así decirlo, «del arte que no está claro que sea arte». Habiéndose «descubierto» hacia fines del siglo xix el arte rupestre paleolítico, una de las sorpresas antropológicas que no dejaron de sucederse por aquel tiempo, no es extraño que se buscara un cobijo para conservar, estudiar y clasificar toda una serie de extraños objetos, cuya belleza formal en ocasiones les hacía acreedores de ser considerados como artísticos, pero de los que se desconocía por completo su intención original; en suma, que en los museos arqueológicos se emplazó todo aquello que despertaba interés por su antigüedad más que por su calidad artística. Por su parte, los de arte contemporáneo, siendo cada vez más éste un arte en constante redefinición, generaban un paréntesis temporal en espera de que en un futuro pudieran integrarse o no dentro de los históricos. En cualquier caso, la primera afirmación «positiva» de un museo de arte contemporáneo fue la que estableció, en 1928, el Museo de Arte Moderno de Nueva York, fruto de la mente brillante de un joven historiador y crítico de arte, Alfred H. Barr, y dos audaces mecenas que sufragaron esta aventura de carácter, no puede obviarse, «privado». Para Barr la cuestión no era ya dedicar un espacio a un arte rechazado por su rareza, sino a un arte con personalidad propia que estaba generando una historia singular. Sin restar el indudable mérito debido a esta empresa, no hay que olvidar que tuvo lugar en plena etapa de entreguerras, cuando, por primera vez, entró en crisis la idea de vanguardia, pues no en vano había consumado sus dos etapas de agitación moderna antes descritas y, además, entre otras cosas porque, también por primera vez, el arte de vanguardia despertó un interés social, aunque todavía restringido a una élites. El que, por ejemplo, Picasso, ya considerado tras la Primera Guerra Mundial como el fascinante y temible epitome de la modernidad artística, no sólo alcanzase una proyección internacional más allá de los foros artísticos y, en conse-

cuencia, que obtuviese rendimientos económicos dignos de un multimillonario ya durante esa década de 1920, nos puede servir de elocuente ilustración al respecto.

De todas formas, entre 1930 y 1960, el arte contemporáneo, aun habiendo generado ya un cada vez más importante mercado y una cada vez mayor credibilidad social, no superó la cota de los círculos minoritarios, con lo que tampoco se convirtió en un producto de consumo masivo, como ocurrió a partir de 1970 hasta la actualidad. Significativamente, durante aproximadamente los últimos veinticinco años, no sólo se han multiplicado de manera abrumadora los museos de arte contemporáneo, sino todo tipo de plataformas institucionales y comerciales para el apoyo, no ya del arte de nuestra época ni del siglo xx, sino, cada vez más, de lo «último». Esta ansiedad polarizada en la novedad no sólo representa, en efecto, el triunfo de lo moderno, sino también el replanteamiento del museo de arte contemporáneo, que ha superado sus dos fases de temporal guarida protectora de un producto en discusión y la de afirmación exultante.

Pero ¿qué problema puede crear un arte contemporáneo apreciado masivamente y reconocido desde todas las instancias para una institución como el museo? Aunque hay quien cree que el problema es meramente funcional y, por tanto, se ciñe al hecho de que el nuevo «tipo» de obras, que emplean soportes y materiales diferentes de los tradicionales, exigen un nuevo modelo de edificio y unos servicios complementarios, hay que estar ciego para no reparar en el calado estético del asunto. Un museo es un archivo cultural que no sólo ordena la memoria del pasado, sino que periódicamente ingresa los productos nuevos del presente en esa secuencia. Así sucedió durante el primer siglo de existencia de los museos públicos, que, como se ha comentado, sólo a fines del xix interrumpieron provisionalmente esta práctica para calmar la inquietud social. Durante el xx, acabó cuajando la idea de que era mejor mantener separado este ámbito de lo contemporáneo, cuya generación de novedades era lo suficientemente amplia y compleja como para formar una historia propia. No obstante, de un tiempo a esta parte, o, como hemos dicho, desde aproximadamente la década de 1970, la creciente polarización en exclusiva hacia lo último, planteó la necesidad de reciclar cualquier pasado que, como tal, resultaba un lastre insoportable. De esta manera, en una primera etapa, hemos comprobado que muchos de los «nuevos museos» de arte contemporáneo decidían, primero, prescindir de la primera mitad del siglo xx, y, luego, iban recortando décadas de arte de la segunda mitad hasta prácticamente convertirse en provisional lecho donde yace la moda o lo nuevo mientras lo son.

Pero ¿qué significa que un museo sólo sea la sede de la cambiante novedad, como si se tratase de una plataforma de exposiciones temporales en exclusiva, al modo de lo que se estuvo llamando estos años una *kunsthalle*? Y si es así, ¿por qué seguir llamándoles museos? A esta

segunda pregunta se puede responder con una obviedad: sea cual sea la interpretación que se haga de lo que es y debe ser un museo, si lo que ahora se entiende como tal no tiene nada que ver con lo que hasta el presente ha definido a este tipo de institución, sería mejor que se buscara una nueva denominación que despejara el equívoco. No hay que olvidar que el arte durante casi toda su larga historia ha vivido al margen de los museos e incluso de cualquier cosa que pudiera parecerseles, así como que los llamados museos públicos, como ya se ha señalado, tienen apenas un par de siglos de historia, con lo que la desaparición de éstos, de llegar a producirse, no debería ser planteada como ninguna tragedia. Por otra parte, aunque la ansiosa deriva hacia lo último, que se insinúa en los «nuevos museos» o «museos de la última generación», entre progresivamente en contradicción con el concepto tradicional de museo como archivo histórico del pasado o del constante ingreso del presente en el pasado, tampoco, a mi juicio, hay que rasgarse las vestiduras por ello, porque los tales, de continuar así, no sólo pueden usar, en efecto, un nuevo y más adecuado nombre para su función, sino formular ésta de otra manera como, por ejemplo, se me ocurre, el lugar artístico donde se recogen novedades que puntualmente se creen artísticas, lo cual es lo mismo que decir: un museo, él mismo, como obra de arte total y no en función de lo que circunstancialmente en él se exhibe. En cualquier caso, se piense lo que se quiera al respecto, lo que parece evidente es que este asunto de los museos de arte contemporáneo está, como se dice, de «rabiosa actualidad» y su debate tiene garantizado futuro, que es justo lo que nos ocupa en el presente ensayo.

Por lo demás, creo que ha llegado el momento de recapitular sobre lo que se ha escrito hasta aquí, porque, tratando de conjeturar acerca de lo que creemos que será el futuro del arte, no sólo es obligado abordar, en primer término, si acaso el arte tiene futuro, lo cual por el momento es tan incierto como indemostrable, sino, en segundo, para dar cierta solidez a nuestras conjeturas, centrar nuestra reflexión en lo único que sabemos: qué es lo que ha sido hasta ahora el arte —su pasado— y, todo lo más, qué es lo que creemos que está pasando ahora, que no es fácil averiguarlo con un simple vistazo, porque lo que verdaderamente está pasando ahora no se parece ni por asomo a lo que se publicita como «actualidad», que es sólo lo que avanta hoy el poder mediático más operativo; es decir, la «actualidad» es tan sólo lo mediaticamente publicitado o, sin más, la publicidad.

En todo caso, ninguna de estas incertidumbres debe descorazonarnos, porque no saber a ciencia cierta todo de algo es característico del parcial o limitado conocimiento humano, sea cual sea su método o su contenido. Quiero decir que lo revolucionario de nuestra época no se ciñe en absoluto a los constrictivos límites de lo que llamábamos o llamamos arte, ni a los muchos más amplios y difusos de lo que queremos embutir en el terreno cultural, ni tan

siquiera los que afectan a la ciencia, que consideramos con dogmática ingenuidad que es el único asidero fiable para nosotros hoy como exploradores adelantados del caos, que es toda la inmensidad abrumadora de lo que no hemos sido capaces todavía de discernir y, por tanto, mínimamente constatar su presencia real. ¿Por qué entonces hay que elegir una pequeña, casi insignificante parte de este aterrador «vacío» de conocimientos para cebarse en que no hay en ella tampoco una certidumbre sobre nada? Por otra parte, a diferencia de los demás conocimientos y experiencias más revalidados hoy socialmente, el arte, ayer u hoy, jamás se ha arrogado la facultad de darnos respuestas, sino, en todo caso, la de hacernos preguntas a través del testimonio de lo que hemos sido y somos sin ser todavía capaces de aprehender su sentido, de manera que se puede afirmar que el arte ha sido y es el «memorizador» de aquellos reflejos que irradiamos sin saber a ciencia cierta por qué y cómo, pero sobre los que tenemos la intuición de que su pérdida –su olvido– nos arrebataría a nosotros y a nuestros descendientes claves esenciales.

Desde la perspectiva que acabo de sugerir se atenúan nuestras ansiedades mal proyectadas sobre la «responsabilidad» por ser incapaz de orientar nuestra fatal irresponsabilidad. Me parece, sin duda, más fértil la vía antitética: sumarnos a las preguntas que el arte actual nos plantea, ahondando en ellas en vez de descalificarlas o trivializarlas. En este sentido, a la legítima inquietud que nos produce hoy no poder abarcar, ni por asomo, los límites del territorio de lo que seguimos nombrando como arte, hay que añadir la todavía más vasta y compleja de la cada vez más indiscriminada extensión social de los «consumidores» de mismo, ese «público» que acoge en su seno a todo el que quiere

entrar sin preguntarle previamente el calado de su saber previo al respecto, de su afición ni de nada. Parece claro, no obstante, que la desbordante extensión del horizonte del arte contemporáneo es simétrica a la de la grey de sus seguidores, con lo que, por lo mismo, las revelaciones de aquél, de producirse, también recaerán sobre muchos más.

Pero ¿qué tiene que ver todo esto con lo que periódicamente se publicita como lo más nuevo entre lo nuevo, atribuyéndolo a banalidades como lo insólito de su soporte material, de su renovado instrumental técnico o cualquiera de sus reformulaciones simbólicas? No creo, por poner ejemplos, que los interrogantes que se hace un vídeo-artista actual o cualquier miembro de esa tribu de los llamados «inmaterialistas» sean diferentes de los que se planteó Francisco de Goya, Miguel Ángel o Fidias, y no lo creo, entre otras cosas, porque aún no hemos acertado a darnos una respuesta pausable del porqué del arte rupestre paleolítico, aunque desde la cuevas de Altamira a Damien Hirst segregamos unas luminosas radiaciones que sentimos que nos son imprescindibles para conocernos y vivirnos más y mejor. De manera que ¿cuál es, en efecto, el futuro del arte? Una buena manera de afrontar esta pregunta sin respuesta es volver a considerar todo lo que el arte ha sido hasta hoy, que por igual forma parte de su pasado. Quizá eso nada tendrá que ver con lo de mañana pero, para conocer el mañana, no existe otra fórmula definitiva que llegar vivo a él y vivirlo. ¿Cómo será, por tanto, el arte del futuro y el futuro del arte? Pues, permítaseme la perogrullada de responder que «lo iremos averiguando», y la averiguación no es sino constatar cómo, cada vez, se nos «presenta» el futuro, o, lo que es lo mismo, cómo se manifiesta como presente.

la arquitectura del nuevo siglo. una vuelta al mundo en diez etapas

LUIS FERNÁNDEZ-GALIANO

Técnica y arte, la arquitectura es también expresión construida de una sociedad. Como técnica lindante con la ingeniería, ha experimentado el impacto de los nuevos materiales y de la innovación en el terreno de la construcción, las estructuras o las instalaciones, enfrentándose al desafío histórico de la sostenibilidad; como arte público, ha sido partícipe —y en ocasiones protagonista— de la renovación del lenguaje visual y las mutaciones estéticas de una época marcada por el espectáculo; como sociología construida, por último, ha dado forma a la colosal transformación que ha urbanizado el planeta reemplazando los paisajes tradicionales por megápolis insomnes.

Los tratados clásicos de los que se ha alimentado la arquitectura de Occidente ya aludían a estas tres dimensiones complementarias cuando teorizaban sobre una disciplina que tanto se solapa con otras fronterizas. Desde el romano Vitruvio, a la arquitectura se le ha asignado la tarea de reconciliar la técnica y el arte con el uso social de sus espacios, y la divisa *firmitas, utilitas, venustas* (solidez, utilidad, belleza) ha sido el resumen taquigráfico de ese propósito. Sin embargo, las tres facetas están tan enrevesadamente trenzadas en las obras concretas de arquitectura que hacen difícil comentarlas de forma separada, y aquí se ha elegido otra estrategia argumental.

En lugar de describir las innovaciones técnicas funcionales y formales que caracterizan la arquitectura de estos

inicios del siglo **xxi**, se ha preferido elegir una decena de episodios en diferentes ciudades del planeta que suministran a la vez una secuencia de realizaciones significativas de las últimas dos décadas, y una ilustración de tendencias o fenómenos de carácter más general. Los episodios, que se presentan en un orden aproximadamente cronológico —desde el Berlín posterior a la caída del Muro, el Bilbao del Guggenheim o la Nueva York del 11-S hasta el Pekín olímpico y las obras titánicas de las autocracias petroleras del Golfo o Rusia— se organizan también de manera que el recorrido se asemeja al de un viaje con escalas alrededor del planeta.

Siempre hacia el ocaso, y siempre en el hemisferio norte —lo que deja no pocas geografías en los márgenes— la ruta se inicia en la Europa que clausuró el siglo **xx** y la Guerra Fría con la demolición de una frontera urbana, transcurre por la América que vio en el derribo de las Torres Gemelas el pistoletazo para el comienzo de una «guerra contra el terror», recorre el Asia que levanta enérgicos signos contruidos de su pujanza económica y se remata en una Rusia encabalgada entre continentes y que afirma —también con la arquitectura— su recuperación tras el trauma de la disolución de la URSS. Al cabo de diez etapas, el círculo se cierra con otra glaciación política, que se suma al enfriamiento económico y a las convulsiones financieras y sociales, en un cúmulo de fracturas y temblores que la arquitectura registra con su exacta aguja de sismógrafo.

Página anterior:

El museo Guggenheim Bilbao,
obra de Frank Gehry.

Berlín sin Muro: la arquitectura de la memoria ante las pugnas ideológicas

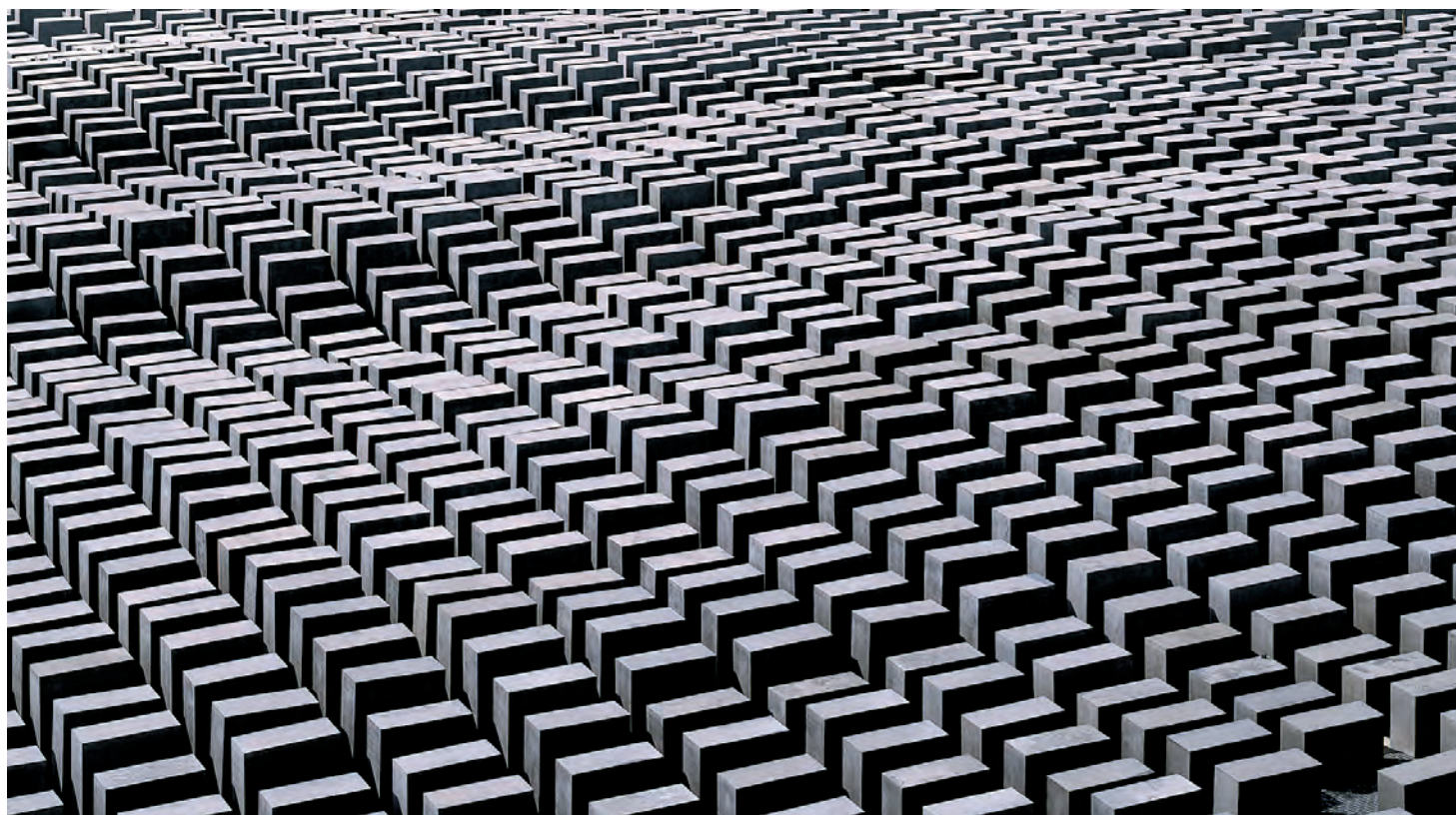
El periplo se inicia en la ciudad del mundo donde la arquitectura más fielmente refleja las ideas. Capital de un imperio totalitario derrotado en 1945, y frontera durante cuatro décadas entre la democracia occidental y el bloque comunista, el Berlín posterior al derrumbamiento del Muro en 1989 ha seguido siendo un laboratorio urbano donde la arquitectura se somete al exigente filtro de la ideología y la memoria. Así ocurre en el Museo Judío del norteamericano de origen polaco Daniel Libeskind, un conjunto de volúmenes fracturados e inestables que se añade a un edificio barroco; en el nuevo Reichstag del británico Norman Foster, una restauración crítica que transforma el carácter de una sede mítica; y en el Memorial del Holocausto del neoyorquino Peter Eisenman, una extensión alabeada de estelas de hormigón que convierte el monumento en un paisaje urbano.

La forma zigzagueante del Museo Judío aludía a los dramáticos cambios de dirección de la historia alemana y a la trágica interrupción de la presencia judía en la ciudad, pero tuvo también una singular importancia arquitectónica. Berlín había sido con la IBA de 1985 —una exposición cuyos objetos eran edificios realizados a escala 1:1 en distintos barrios de la ciudad— el principal escenario del movimiento posmoderno, que preconizaba el retorno hacia las arquitecturas clasicistas por oposición a las abstracciones de la

modernidad; con el proyecto de Libeskind, que se hizo público por las fechas de la caída del Muro, Berlín construiría un icono de la deconstrucción, una tendencia rival —lanzada con una exposición en el MoMA de Nueva York durante el verano de 1988— que defendía arquitecturas fracturadas y catastróficas como expresión de un mundo convulso.

Para convulsiones, desde luego, las provocadas por las dos guerras mundiales que tuvieron en Berlín su epicentro, y que dejaron en la ciudad las ruinas del antiguo parlamento como testigo mudo del colapso de la democracia y del wagneriano derrumbamiento del expansionismo germano. Cuando, tras la capitulación de Gorbachov frente a Reagan y Thatcher, el final de la Guerra Fría permitió la reunificación alemana y el retorno de la capitalidad a Berlín, Foster rehabilitó el viejo Reichstag como nuevo Parlamento de una nación decidida a impedir el regreso de los fantasmas de un pasado ominoso: para ello, coronó la mole guillermina con una cúpula de vidrio animada por rampas helicoidales que sirve como mirador urbano y sitúa simbólicamente a la ciudadanía por encima de sus representantes políticos, supervisando su asamblea para impedir nuevos descarrilamientos históricos.

Próximo a este Reichstag regenerado —que incluso el artista Christo exorcizó cubriéndolo con lonas antes de comenzar las obras— Eisenman construyó un colosal y lírico memorial del exterminio judío con una malla de prismas de hormigón que es a la vez paisaje ondulante de



Memorial del Holocausto en Berlín, obra de Peter Eisenman.



La Tate Modern de Londres, obra de Herzog de Meuron.

campos de cultivo y laberinto de desasosiego entre tumbas exactas. Concebido en sus inicios con el escultor Richard Serra, el monumento conmemorativo —tan diferente en su áspera abstracción de la mayor parte de los museos del Holocausto que han florecido en tiempos recientes— es un gesto penitencial en el corazón del horror, y al mismo tiempo un eficaz ejemplo de la capacidad de la arquitectura para expresar ideas a través de las formas.

Rotterdam o Basilea: nuevos paisajes y viejas ciudades en una Europa indecisa

Tras el agotamiento del estilo posmoderno —que sí tuvo a Berlín de escenario coral y se promovió intelectualmente desde Milán y Nueva York— el debate de la arquitectura no se desplazó al París de los grandes proyectos presidenciales de Mitterrand, con su combinación de monumentalidad geométrica y el *glamour* de las celebridades, ni al Londres de Blair, que festejó el milenio con una tercera vía arquitectónica, tecnológica y *cool*. Fue a recaer en dos ciudades europeas de tamaño intermedio, la holandesa Rotterdam y la suiza Basilea. En la primera, numerosos arquitectos jóvenes, inspirados por el talento abrasivo de Rem Koolhaas, y entre los cuales deben destacarse los agrupados bajo las siglas MVRDV, exacerbaron el lenguaje moderno con acentos de las utopías constructivistas rusas de los años veinte y lo extendieron al paisaje urbano; en la segunda, una nueva generación de suizos alemanes, don-

de enseguida destacó la energía creativa de Jacques Herzog y Pierre de Meuron —con el permanente contrapunto rural y esencialista de Peter Zumthor— crearon un reducto de excelencia constructiva, exigencia artística y sensibilidad hacia la herencia material de territorios ancestrales.

La hipermodernidad holandesa se alimentaba de la *tabula rasa* de una ciudad devastada por la guerra sobre el territorio artificial de un país de *polders*, pero también de la fascinación futurista de Koolhaas por la congestión metropolitana de una Nueva York que fue durante años su ciudad de adopción, objeto de estudio y laboratorio intelectual, esto último en las redomas del IAUS (Institute for Architecture and Urban Studies) dirigido por Peter Eisenman. Combinando la gramática formal de Le Corbusier con las audacias diagonales rusas y el pragmatismo americano, en los Países Bajos se gestó una escuela optimista y juguetona, que no dudó en coquetear con los quiebros y alabeos del deconstructivismo anglosajón —alimentado en no pequeña medida por la extrema ductilidad que ofrecen los nuevos sistemas de representación informática—, pero que halló su mejor manifestación en un artificioso paisajismo que penetra surrealmente en los edificios con la topografía del entorno, para crear una «sección libre» que da una vuelta de tuerca a la «planta libre» de las vanguardias históricas.

En contraste, los suizos alemanes desarrollaron un «grado cero» de la arquitectura, a través de prismas elementales exquisitamente contruidos, profundamente

arraigados en las tradiciones y el territorio del país alpino, pero influidos también por la enseñanza rigorista del Aldo Rossi que fue profesor de Herzog de Meuron en la ETH de Zúrich. Desafiantemente arcaicos, y al mismo tiempo estrechamente vinculados con la escena artística —inicialmente a través de Joseph Beuys, y después mediante múltiples colaboradores del mundo del arte—, los dos socios de Basilea se erigieron en líderes de su generación con una secuencia de cajas ornamentadas de gran refinamiento material y táctil, y con una serie de intervenciones en edificios industriales —muy singularmente su conversión de una central eléctrica en la nueva sede de la Tate Gallery londinense— que mostraron la vigencia de una arquitectura de la continuidad.

Ante una Europa próspera y políticamente fatigada, indecisa entre el mesianismo moderno de la construcción *ex novo* de la ciudad contemporánea y el vínculo —cultural y emocional— con su heterogéneo patrimonio urbano, holandeses y suizos suministraron modelos arquitectónicos y urbanísticos enfrentados, estableciéndose un fértil diálogo disciplinar entre Rotterdam y Basilea que con el tiempo llegaría a provocar una cauta convergencia entre las dos escuelas.

Bilbao y el Guggenheim: el espectáculo del museo como motor urbano

En 1997, la inauguración de la sede bilbaína del Museo Guggenheim —una ondulante acumulación escultórica de chapas de titanio diseñada por el californiano Frank Gehry— fue un acontecimiento mediático que alteró el curso de la arquitectura y los museos. Desde luego, la institución neoyorquina contaba ya con una sede original de gran singularidad y belleza arquitectónica —la famosa rampa helicoidal levantada por Frank Lloyd Wright en la Quinta Avenida—, y su emblemático edificio junto a la ría de Bilbao había tenido precedentes icónicos tan significativos como la ópera de Sidney, donde el danés Jørn Utzon construyó unas velas de hormigón que se convirtieron en el símbolo de Australia, o —ya en el terreno de los museos— el Pompidou parisino, donde el italiano Renzo Piano y el británico Richard Rogers interpretaron con su futurismo alegre, colorista y tecnológico el espíritu contracultural de los jóvenes de las revueltas del 68.

El Guggenheim bilbaíno daba un paso más, porque subordinaba el arte por entero al espectáculo de la arquitectura, y hacía de ésta una escultura gigantesca, delicadamente mate en sus reflejos y alborotada en su tormentoso movimiento detenido. Éxito de crítica y de público, el museo atrajo numerosos visitantes a una áspera ciudad de industria obsoleta hasta entonces apartada de los circuitos artísticos o turísticos, se convirtió en un poderoso motor de regeneración urbana y mostró la capacidad de las infraestructuras culturales para contribuir al tránsito hacia una economía de servicios. Lo que en España se llamó «efecto Guggenheim» y fuera del país «efec-

to Bilbao» se extendió como un incendio en un secarral, y cada alcalde de una ciudad en decadencia procuró dotarse de un edificio emblemático que llamase la atención de turistas e inversores, mejorase la autoestima y sirviese de logo para un cambio de imagen.

Este uso de la arquitectura para la modernización identitaria y el *rebranding* urbano —que llegó a afectar a metrópolis de la dimensión y el carácter de Londres o Roma— acentuó la deriva disciplinar hacia la diferenciación escultórica, ya que cada nueva sede cultural o deportiva debía ser inconfundible y sorprendente: los museos desde luego, pero también las bibliotecas, los auditorios o los estadios tenían que reconciliar sus funciones propias con su papel simbólico, y aun edificios tan exigentes en su organización como las estaciones o los aeropuertos —en el propio Bilbao, las estaciones de metro las realizó Norman Foster, y el aeropuerto Santiago Calatrava— se pusieron al servicio de la identidad urbana, en una senda ya recorrida por las grandes corporaciones que promueven rascacielos singulares como imagen de marca en el perfil coral de la ciudad.

Guy Debord teorizó «la sociedad del espectáculo» en 1967, pero cuatro décadas después sus intuiciones siguen plenamente vigentes. La absorción de la arquitectura por el mundo del espectáculo tiene un sabor agríndice: por un lado, otorga una mayor visibilidad a las obras, y hace a éstas objeto del debate social en los medios, como ha podido verse en realizaciones recientes de maestros tan exigentes y secretos como Álvaro Siza o Rafael Moneo; por otro, transforma a los arquitectos en celebridades del *glamour* y la moda, y si Gehry diseña joyas para Tiffany's y Koolhaas o Herzog de Meuron proyectan las tiendas de Prada, la angloiraquí Zaha Hadid levanta para Chanel un pabellón alabeado y portátil, y todos ellos aparecen con frecuencia en la publicidad de consumo de lujo como representantes excelsos de la discriminación estética y la elegancia vanguardista.

Nueva York tras el 11-S: el futuro del rascacielos y el futuro del imperio

La cuarta etapa de nuestro viaje alrededor del mundo deja la Europa donde tantas expectativas había despertado el final de la Guerra Fría y el disfrute hedonista de los dividendos de la paz, y cruzando el Atlántico hace escala en Nueva York, teatro de un titánico atentado que produjo una trágica masacre y torció el rumbo de la historia contemporánea. El grupo suicida de jóvenes militantes islámicos dirigido por el arquitecto y urbanista Mohamed Atta derribó en Manhattan dos torres diseñadas por el americano de origen japonés Minoru Yamasaki que simbolizaban el poder financiero de la ciudad y el liderazgo global de Estados Unidos, provocando con su atroz acción una crisis geopolítica sin precedentes, y poniendo de paso en cuestión el futuro del rascacielos, el edificio que mejor representa los desafíos arquitectónicos del siglo xx.

Página siguiente:

Sede del *New York Times* en Nueva York, obra de Renzo Piano.

En efecto, la destrucción de las Torres Gemelas redibujó en el planeta las fronteras del conflicto, y la rivalidad ya extinta entre capitalismo y comunismo fue reemplazada por el enfrentamiento entre Occidente y el fundamentalismo islámico; al mismo tiempo, el prestigio de la superpotencia dirigida erráticamente por George W. Bush sufrió un golpe devastador —agudizado por los errores de las posteriores «guerras contra el terror» en Afganistán e Irak—, su economía experimentó el lastre del gasto militar y el barroquismo financiero, y la metrópoli neoyorquina vio abrirse una herida que todavía no ha cicatrizado: el fiasco intelectual, estético y administrativo de los concursos de arquitectura convocados para regenerar el vacío ominoso de la Zona Cero es uno más de los signos de una pérdida de pulso que hace temer una decadencia anunciada.

Sin embargo, el pronosticado ocaso del rascacielos —que a su complejidad y a su coste añadía ahora una extrema vulnerabilidad— ha estado muy lejos de producirse, y las torres han seguido levantándose por doquier. Con una seguridad revisada, y los inevitables incrementos en sus presupuestos, los grandes protagonistas públicos y privados del poder han continuado construyendo rascacielos para manifestar su pujanza mediante la altura: aunque muchas corporaciones han vuelto los ojos hacia los parques de oficinas, y aunque más allá de los 200 metros las torres apenas tiene justificación económica, la pugna por los récords planetarios o regionales sigue alimentando la competencia entre ciudades o países, obteniendo la atención de los medios y suscitando la curiosidad de las gentes.

La propia Nueva York que sufrió el golpe devastador del 11 de septiembre no renuncia a seguir siendo conocida como «la ciudad de los rascacielos», y persevera en la construcción y el proyecto de nuevas torres, a menudo vinculadas con su persistente liderazgo cultural y artístico, como las sedes de los grupos periodísticos Hearst y New York Times (diseñadas, respectivamente, por Norman Foster y Renzo Piano), el pequeño apilamiento del New Museum (obra de los japoneses Sejima y Nishizawa) o el rascacielos residencial proyectado por el francés Jean Nouvel junto al MoMA: un sector este —el de las viviendas de lujo firmadas por grandes arquitectos— que si ha prosperado en Manhattan ha dado lugar en Chicago, cuna del rascacielos y ciudad que en su mitología arquitectónica une a Sullivan y Wright con Mies van der Rohe, a un espectacular proyecto en altura de Santiago Calatrava, el mismo arquitecto español que está construyendo, con su catedralicia estación de metro, la única obra relevante de la atribulada Zona Cero neoyorquina.

Las Vegas como paradigma: el urbanismo del ocio y la tematización del mundo

América alumbró los rascacielos, que llevan la densidad urbana a su extremo más hiperbólico; pero también dio carta de naturaleza al urbanismo más disperso, que con ayuda del automóvil extiende la ciudad sobre el territorio como una delgada alfombra de casas y jardines. Tal subur-

banización unánime, que malgasta el espacio y el tiempo despilfarrando materiales, agua, energía y terreno —amén de infraestructuras de transporte—, y que se ha exportado con gran éxito al resto del mundo, relega el dominio colectivo a grandes aglomeraciones comerciales que a menudo se presentan con el ropaje de la urbanidad tradicional, interpretada figurativamente con los mismos recursos escenográficos —admirados por la retina pop de Warhol en el terreno del arte, y por la de Venturi y Scott-Brown en el campo de la arquitectura— que los parques de atracciones de Disney o los casinos temáticos de Las Vegas.

Esta ciudad de Nevada, que es también la que experimenta el crecimiento más rápido de Estados Unidos, sirve bien de paradigma de la urbanización posmoderna, cuyas tendencias exagera hasta el paroxismo, y suministra una eficaz metáfora del auge contemporáneo del «capitalismo de casino», tan brillante, ruidoso y masivo como las salas de juego que extienden sin solución de continuidad los vestíbulos de los interminables hoteles al servicio del espejismo del ocio en esta urbe de neón. La tematización egipcia o veneciana de los locales de Las Vegas —como los pueblos del Lejano Oeste o los castillos de Blancanieves en las innumerables disneylandias dispersas por el mundo— rebota como un eco testarudo en cada *mall* de América y del mundo, y el urbanismo de consumo remeda con torpeza las trazas de una urbanidad desvanecida.

La centralidad del comercio en estas nuevas formas de ocupar el territorio —admirablemente analizada por Koolhaas en su descripción del *junkspace*, el «espacio basura» contemporáneo— resulta indiscutida, y la morfología del centro comercial —grandes superficies de venta y *food courts* incluidas— se infiltra en las restantes infraestructuras de transporte, ocio, deporte, cultura, salud o trabajo que articulan con sus nódulos de actividad la extensión indiscriminada de la construcción residencial: los aeropuertos y las estaciones, los parques de atracciones, los estadios, los museos y aun los hospitales o los campus de enseñanza, investigación y negocios sufren la penetración invasiva del *mall*, que con sus comercios y restaurantes complementa y a la larga protagoniza los ámbitos de relación y encuentro en la suburbanización temática del mundo.

Incluso las ciudades compactas de la tradición europea, ampliadas con periferias anónimas e indistintas de baja densidad, reformulan sus centros históricos como espacios de ocio y turismo, amplios centros comerciales sin techo en los que se combinan las boutiques, tiendas de moda, bares y terrazas con el ocasional palacio, iglesia o museo. Así, ciudades como la Barcelona de Bohigas y Miralles, que se propuso en el escaparate de los Juegos Olímpicos de 1992 como un ejemplar modelo de transformación urbana, atento tanto a «higienizar el casco» como a «monumentalizar la periferia», encuentra que la deriva contemporánea ha creado un ámbito ciudadano al servicio de los visitantes ocasionales, muy alejado de los fundamentos modernos, vanguardistas y aun utópicos del proyecto inicial.

Página anterior:

El museo Guggenheim Las Vegas, obra de OMA/Rem Koolhaas.

Tokio en dibujos animados: tradición y modernidad en la densidad japonesa

El ecuador de nuestro viaje es más bien el meridiano que señala el cambio de fecha. Al otro lado del océano Pacífico, la sexta escala del periplo tiene lugar en Tokio, una metrópoli formalmente desmemoriada donde la pervivencia de los hábitos tradicionales coexiste con un paisaje urbano futurista, abigarrado de signos y con la animación espasmódica de los dibujos animados. A principios del siglo pasado, la fascinación por un «japonesismo» exótico coloreó el lenguaje de las vanguardias artísticas, mientras la modernidad arquitectónica importó del «imperio del sol» la racionalidad extrema de la construcción en madera, la ligereza modular de las casas divididas con *tatamis* y papel de arroz, y el laconismo depurado y ceremonioso de los objetos: desde Frank Lloyd Wright y sus discípulos vieneses en California, hasta el viaje de ida y vuelta del berlinés Bruno Taut o el descubrimiento del Lejano Oriente por parte de Alvar Aalto y sus colegas del organicismo escandinavo, Japón y modernidad han sido sinónimos arquitectónicos.

Hoy, sin embargo, la hiperurbanidad japonesa suministra un modelo muy alejado del ensamblamiento en penumbra de la casa intemporal. Si volviese a escribirlo, el *Elogio de la sombra* de Tanizaki —que tanto ha alimentado la sensibilidad zen del minimalismo occidental— sería en nuestros tiempos un «elogio del neón», exponente emblemático de una cultura pop, juvenil y ultracomercial, tan estrepitosa como la de Las Vegas, aunque adornada aquí del infantilismo de los *manga* y el autismo cibernético de los *otaku*, y al cabo entregada a la veneración de las marcas de lujo, que jalonan el paisaje de la ciudad con sus sedes exquisitas y herméticas.

Más allá de los jardines immaculados o los museos de geometría exacta —como tantos de los que ha ejecutado en hormigón y vidrio Tadao Ando, reuniendo felizmente los lenguajes formales de Le Corbusier y Louis Kahn—, son las tiendas de moda el mejor escaparate de la temperatura social del Japón actual. Realizadas en ocasiones por arquitectos extranjeros —como el extraordinario cristal facetado diseñado para Prada por Herzog de Meuron, o el lírico prisma translúcido levantado para Hermés por Renzo Piano— pero con frecuencia exponentes de la más refinada arquitectura local —de la superposición azarosa de Sejima y Nishizawa para Dior a las celosías arborescentes de Toyo Ito para Tod's— las sedes de las firmas de lujo en Omotesando o en Ginza, los dos barrios de Tokio donde se concentra el mundo de la moda, son testimonio de una exacerbación hiperbólica del consumo suntuario que desborda ampliamente sus fuentes originales en Europa o América.

Frente a esa ostentación inocente palidecen las grandes infraestructuras del transporte en el resto del país —que tienen sin embargo ejemplos tan destacados como el colosal aeropuerto de Osaka, construido por Piano sobre una isla artificial, o la delicada terminal marítima de Yokohama, realizada por Zaera y Moussavi con plataformas alabeadas

de madera— o las más singulares obras culturales —entre las cuales la mediateca de Sendai, sostenida por Ito con haces enredados de pilares metálicos, o el museo de Kanazawa, delimitado por Sejima y Nishizawa con un evanescente perímetro circular—, todas las cuales se inscriben en un dominio público fluido y liviano, reflectante y carenado, tan límpido como frígido, y en cualquier caso carente de la magia magnética y centrípeta de esos reductos privados y exclusivos del lujo más sofisticado y más vacío en el corazón de ese «imperio de los signos» que es la ciudad de Tokio.

Pekín olímpico: los iconos del protagonismo chino en el auge asiático

Frente al Tokio de la moda, el Pekín del espectáculo. La inauguración y el desarrollo de los Juegos Olímpicos durante el verano de 2008 permitieron a China enorgullecerse de sus logros económicos y sociales, presentando al mundo un formidable ejemplo de su capacidad organizativa con un evento donde la arquitectura fue algo más que un mero escenario mudo de las ceremonias y las competiciones. Desde la nueva terminal del aeropuerto por el que llegaron atletas, espectadores o periodistas, hasta la sede de la televisión que transmitió los Juegos, pasando por los propios recintos deportivos, encabezados por el estadio y las piscinas, las grandes obras realizadas para el acontecimiento —pese a estar diseñadas casi todas por arquitectos extranjeros— evidenciaron la ambición de excelencia de China, y dieron a la vez testimonio del camino recorrido por el «imperio del centro» durante los treinta años transcurridos desde la mutación política de 1978, cuando el maoísmo caótico de la Revolución Cultural fue reemplazado por el capitalismo de partido único impulsado por Deng Xiaoping.

La nueva terminal, que es además el edificio más grande del planeta, fue realizada por Norman Foster —autor igualmente del aeropuerto de Hong Kong, lo mismo que el de Osaka sobre una isla artificial— con el característico refinamiento tecnológico de la oficina británica, que supo interpretar las columnas rojas o las cubiertas flotantes de la construcción tradicional con el acero y el vidrio de la alta ingeniería, para crear un recinto interminable y luminoso que protege a los pasajeros de los aviones bajo un techo tan liviano como un dragón de fiesta o un cometa de papel. La instalación se inauguró un año antes de los Juegos, lo mismo que otra gran obra promovida para el evento, el Teatro Nacional levantado por el francés Paul Andreu —curiosamente también arquitecto de aeropuertos— junto a Tiananmen, en forma de una gigantesca cúpula de titanio que emerge sobre el agua quieta de un vasto estanque.

La competición deportiva tuvo un protagonista líquido en las piscinas contenidas en el que muy pronto se conoció como «el cubo de agua», un gran prisma de fachada burbujeante realizada con almohadas translúcidas de un plástico llamado ETFE (etil tetrafluoretileno) por el equipo de Australia PTW; y, sobre todo, disfrutó del formidable escenario del Estadio Olímpico, una titánica madeja de acero imagi-

Página anterior:

Estadio Olímpico de Pekín,
obra de Herzog de Meuron.

ginada por los suizos Herzog de Meuron con la ayuda del artista chino Ai Weiwei, también distinguida por el público con un apodo cariñoso, «el nido de pájaro», y cuya extraordinaria singularidad formal lo ha hecho icono de los Juegos y símbolo de la pujanza china, que llegó al paroxismo con las espectaculares ceremonias de apertura y clausura, realizadas tanto por la coreografía y la pirotecnia como por la imagen espectral y emocionante del nido nocturno.

Inevitablemente, la sede de la televisión —dos torres enlazadas por sus cabezas para conformar el marco doblado de una colosal puerta urbana, diseñadas por el holandés Rem Koolhaas— fue un edificio más polémico, y no tanto porque no llegase a terminarse a tiempo para los Juegos como porque el carácter estatal de la información es uno de los rasgos más polémicos del país, que combina el éxito económico con un dirigismo social más estricto que en Occidente.

Astana en las estepas: una nueva capital en el territorio del Gran Juego

La octava escala es sin duda la más exótica, porque las estepas del Asia central evocan menos los logros arquitectónicos que la música de Borodin o la literatura de Kipling sobre el Gran Juego geoestratégico de los imperios euroasiáticos. Tierras de tránsito y nomadeo, hasta hace bien poco muchos citarían la *yurta* —una tienda circular de exquisita depuración constructiva— como la aportación más original de estas estepas a la historia de la habitación humana. Con la disolución de la Unión Soviética, sin embargo, en la escena internacional apareció un nuevo actor, la república de Kazajistán, dotada de petróleo y de un presidente carismático decidido a dejar huella arquitectónica con una nueva capital: la existente Almaty —la mítica Alma Ata— sería reemplazada por Astana, una ciudad creada *ex novo* sobre la ruta del ferrocarril transibe-

riano, y muchos de los arquitectos más importantes del mundo serían convocados a su construcción.

En la tradición de la Chandigarh del Pandit Nehru o la Brasilia de Juscelino Kubitschek (desarrolladas, respectivamente, por Le Corbusier y por Lucio Costa y Oscar Niemeyer), la Astana del presidente kazajo Nursultán Nazarbayev ha sido trazada por el japonés Kisho Kurokawa, y tiene al británico Norman Foster como autor de sus edificios más significativos. Así, Kazajistán no se reconoce sólo como el país del personaje interpretado por el actor británico Sacha Baron Cohen —el polémico Borat— y Astana no se asocia únicamente a un equipo ciclista: el país y su nueva capital han ingresado con audacia insólita en el relato de la arquitectura contemporánea.

Desde luego, Foster no es el único occidental con encargos importantes en Kazajistán. Pese al traslado administrativo de la capitalidad, las rentas del petróleo siguen promoviendo un singular *boom* constructivo en la vieja Almaty, donde muchas oficinas norteamericanas y europeas —incluyendo la OMA de Rem Koolhaas, que levanta en las afueras de la ciudad un gran campus tecnológico— expresan en el territorio el vigor económico del país. En Astana, sin embargo, la firma londinense es la protagonista absoluta de la arquitectura emblemática, con una colosal pirámide ya terminada y con una enorme carpa transparente que será el techo de la ciudad cuando se remate.

La pirámide, denominada Palacio de la Paz y la Reconciliación —pero inevitablemente conocida entre el público y los medios como «pirámide de la paz»— es la sede de unos congresos interconfesionales periódicos, y procura conciliar las diferentes razas, culturas y religiones del país a través de su geometría arcaica y exacta, coronada con un vértice translúcido de inocentes vidrieras con palomas. La carpa, que aloja 100.000 metros cuadrados de espacio de ocio bajo una superficie de ETFE sostenida por mástiles y cables, duplica sobradamente la altura de la pirámide, y constituye casi su reverso simbólico, estableciéndose un inesperado diálogo entre las aristas de acero del templo ideológico y los alabeos de plástico de la titánica tienda al servicio del espectáculo y el consumo, reuniendo así las viejas identidades tribales y religiosas con la nueva pertenencia a una tribu global que sólo venera la prosperidad y el entretenimiento.

Dubai y el Golfo: las ciudades del petróleo y el desafío de la sostenibilidad

Nuestra siguiente escala nos lleva a otro *boom* inmobiliario impulsado por el petróleo, pero en este caso de tal dimensión y rapidez que teóricos de la ciudad contemporánea como Rem Koolhaas no han dudado en calificarlo de «una nueva urbanidad», una forma hasta ahora inédita de producir tejido urbano. Lindantes con la ciencia-ficción, las construcciones de los emiratos del Golfo Pérsico —inicialmente alimentadas por la explotación de los pozos, pero cada vez más vinculadas a los flujos financieros y turísticos— se extienden desde



Pirámide de la Paz y de la Reconciliación en Astana, obra de Norman Foster.



Ecociudad proyectada por OMA/Rem Koolhaas en Abu Dhabi.



Ecociudad proyectada por Norman Foster en Ras al Khaimah.

un paisaje surreal de rascacielos que surgen de la arena del desierto hasta un rosario de islas artificiales en forma de continentes o palmeras, e incluyen un sinnúmero de infraestructuras educativas y culturales que alojan franquicias de los principales museos y universidades de Estados Unidos y Europa.

Dubai fue en muchos sentidos pionero, porque al poseer reservas de petróleo muy inferiores a las de otros emiratos se redefinió pronto como un centro financiero regional que pudiera reemplazar en Oriente Medio a un Beirut devastado por la guerra y los conflictos políticos, y como un destino turístico de lujo para los nuevos millonarios de Rusia y Europa. Levantado con la *expertise* de los *project managers* anglosajones y con el esfuerzo de un ejército de trabajadores inmigrantes de India, Paquistán y el sureste asiático sin apenas derechos laborales o civiles, este bosque de rascacielos con una orla de islas temáticas tiene a gala haber culminado, con el Burj al Arab de la firma británica Atkins, el hotel más lujoso del mundo, y con el Burj Dubai de la norteamericana SOM, el edificio más alto del planeta: son récords económicos y téc-

nicos, y sin duda también valiosos indicadores sociales, pero por desgracia dicen poco de la calidad de la arquitectura, donde la acumulación de firmas importantes no ha dejado aún obras magistrales.

Diferente es la estrategia de Qatar, que aspira a convertirse en un centro intelectual a través de una ambiciosa ciudad de la educación diseñada por arquitectos globales como el japonés Arata Isozaki, el mexicano Ricardo Legorreta, el norteamericano de origen argentino César Pelli y los holandeses de OMA, y diferentes son también los objetivos políticos y urbanos de otros dos emiratos: Ras al Khaimah, que intenta promover el turismo sostenible en un enclave de especial belleza natural; y Abu Dhabi, la capital de los Emiratos Árabes Unidos, que ha puesto en marcha un espectacular distrito cultural con sucursales del Guggenheim y el Louvre.

Los proyectos más visionarios de Ras al Khaimah — entre los cuales están un onírico centro turístico en lo alto de las montañas y una ciudad ecológica en la costa, con un emblemático centro de convenciones esférico— son todos del mismo Koolhaas que ha teorizado la explosión urbana del Golfo. En Abu Dhabi, por el contrario, la participación de grandes figuras es más coral, y si Frank Gehry, Jean Nouvel, Zaha Hadid o Tadao Ando se reparten los museos y el teatro del distrito cultural, la poderosa oficina del ubicuo Norman Foster lleva a cabo desde una ejemplar ciudad sostenible (carbón neutral) con transporte colectivo y autosuficiencia energética hasta una lírica interpretación del bazar tradicional en el nuevo Mercado Central de la ciudad. De forma inesperada, en el lugar del mundo de mayores reservas energéticas no se promueve sólo la ostentación y el consumo: como muestran las ecociudades de Koolhaas y Foster, la abundancia no excluye el ensayo de las formas futuras de la austeridad o la escasez.

De Moscú a San Petersburgo: las obras titánicas de la autocracia rusa

El viaje llega a su término muy cerca de donde se inició, en la misma Rusia que situaba en Berlín la frontera física y simbólica de la Guerra Fría, y que estimulada por el control del petróleo y el gas que necesita buena parte de Europa recupera el orgullo imperial de la autocracia zarista y la autoestima implacable del estalinismo soviético. En sintonía con los autoritarismos orientales de Pekín, Astana o Dubai, y disfrutando como esas ciudades del latido impulsivo de una brusca prosperidad, Moscú pone en marcha un turbión de megaproyectos que definen con la elocuencia de la arquitectura las ambiciones renovadas del coloso euroasiático: un *boom* constructivo que, si bien centrado inevitablemente en la capital del país, alcanza a muchas otras ciudades, y muy singularmente a la histórica San Petersburgo.

En las dos urbes tienen una presencia muy significativa los arquitectos británicos, pero en Moscú es obligado subrayar el protagonismo material y mediático del mismo Foster que ha diseñado el aeropuerto de Pekín, la «pirá-



El complejo Crystal Island en Moscú, proyectado por Norman Foster.

mide de la paz» en Astana o la ciudad sostenible de Abu Dhabi: tanto con la Torre Rusia, cuyos 612 metros la convertirán en el rascacielos más alto de Europa, como con la Crystal Island al borde del río Moscova, una auténtica ciudad bajo una descomunal cubierta helicoidal que, además de mejorar su comportamiento climático, hará del conjunto la construcción más grande del planeta —superando a su propia Terminal 3 de Pekín, que por ahora detenta el récord— Foster representa adecuadamente el regenerado vigor del país, que no tolera ya ser tratado, como mostró en la crisis de Georgia, con el conmisericordioso desdén que suscitó la descomposición de la Unión Soviética y el subsecuente declive de la potencia rusa.

Mención aparte merece San Petersburgo, capital cultural y cuna de un Vladimir Putin que ha situado en ella la sede de Gazprom, el gigante energético ruso, para el que la oficina escocesa RMJM —tras un polémico concurso en el que se invitó a participar a las grandes estrellas del panorama internacional— va a levantar un colosal rascacielos que hará diminuta la catedral de Smolny al otro lado del Neva, manifestando con elocuencia el papel de los combustibles fósiles en el renacimiento de Rusia, que intimida a los gobernantes de Europa Oriental a través de los gaseoductos y que se permite el lujo de tener a un ex canciller de Alemania en la nómina de su empresa energética. Al finalizar nuestro trayecto, no sabemos ya si la Guerra Fría terminó de verdad hace dos décadas, pero sí estamos seguros de que la arquitectura continuará dando

expresión a las ambiciones y a los conflictos, a los logros y a las decepciones de los países y de los regímenes, de las empresas y de los pueblos.

Cerrando un círculo más vicioso que virtuoso, es al cabo la oficina londinense cuya remodelación del Reichstag mantuvo los grafiti obscenos escritos en caracteres cirílicos por los soldados rusos que tomaron Berlín —no en vano la más frecuentemente mencionada en este itinerario, donde aparece en siete de sus diez capítulos, y sin duda también la más agresivamente global de todas ellas— la que hoy expresa el poder de Rusia, de China o de los Emiratos con arquitecturas emblemáticas. Se ha completado un ciclo histórico, y el final del mundo bipolar que permitió la reunificación de Alemania tras la caída del Muro en 1989 ha dado lugar —tras el breve intervalo de una única superpotencia que ha fracasado en la gobernanza global— a un escenario multipolar que la arquitectura subraya con la proliferación de sus núcleos de condensación.

Un epílogo provisional: ocaso o aurora de una disciplina en mutación

Es difícil evitar —en este viaje siempre hacia el ocaso— un tono de melancolía en las conclusiones del relato, ya que el itinerario de la arquitectura durante las últimas dos décadas ha transformado una disciplina artesanal y modesta, basada en los conocimientos técnicos, el pragmatismo funcional y la discriminación estética, en una actividad que linda con el estrépito de la publicidad, la avidez del consumo y

el torbellino de la moda. La humildad, la perseverancia y el silencio que solían caracterizarla han sido sustituidos por el aplomo jactancioso, la inventiva caprichosa y una locuaz justificación de propuestas disparatadas que sólo se explican por el apetito inagotable de novedades en las pupilas y el paladar fatigados de una sociedad demasiado próspera.

Los grandes desafíos de una humanidad que ya es mayoritariamente urbana —del cambio climático o la construcción sostenible a la orquestación material de la vida colectiva en megápolis como México, São Paulo, Lagos o Calcuta— parecen serle ajenos a esta práctica ensimismada, extraordinariamente eficaz en la creación de obras emblemáticas o icónicas y dramáticamente incapaz de mejorar de forma significativa la habitabilidad y la belleza de la ciudad contemporánea. Están siendo, como tantas veces se ha dicho, buenos tiempos para la arquitectura (entendida restrictivamente como la construcción de edificios singulares) y malos tiempos para la ciudad, es decir, para ese ámbito que a todos pertenece y a todos representa.

Nunca en la historia reciente han sido los arquitectos tan celebrados, y quizá nunca tampoco han sido tan importantes para conformar el entorno que habitamos. Hace sólo medio siglo, arquitectos anónimos —o conocidos sólo por sus colegas y los especialistas— trazaban en sus estu-

dios planes urbanos y grandes proyectos de habitación colectiva que afectaban decisivamente a la vida cotidiana de la mayoría; hoy, arquitectos convertidos en estrellas mediáticas actúan de árbitros de la moda y dictadores del gusto, pero apenas tienen capacidad para intervenir en las grandes decisiones que modelan la ciudad y el territorio, determinadas casi exclusivamente por los vectores económicos y los flujos de movimiento que cristalizan en las infraestructuras del transporte.

La arquitectura, en todo caso, es una disciplina arcaica y tenaz, que si ha sufrido un proceso desconcertante de mutaciones para acomodarse a la sociedad del espectáculo, no por ello ha abandonado su núcleo esencial de inteligencia técnico-constructiva, orquestación de las cambiantes necesidades sociales y expresión simbólica de la naturaleza de los tiempos: las venerables *firmitas*, *utilitas* y *venustas* vitruvianas. Quizá por ello, el tono elegíaco de estas conclusiones podría ser equívoco, e incompatible con la testaruda confianza que demanda el ejercicio de esta profesión exigente, experta en reconciliar el pesimismo de la inteligencia con el optimismo de la voluntad. Al dar la vuelta al mundo en dirección poniente ganamos un día en el trayecto, y acaso esta luz incierta que tomamos por acaso sea en realidad una aurora de este arte útil y del mundo.

¿fronteras y conocimiento en música? unos apuntes

LUIS DE PABLO

Es tentador pensar que el tema de este trabajo cae fuera de las capacidades –incluso de los intereses– de un artista. En mi caso, de un compositor, lo que no hace sino agravar las cosas, a la vista de la elusiva naturaleza de la música como lenguaje.

Y es incluso posible que así sea. Todo depende del sentido que se dé a la palabra clave: «Conocimiento».

No caeré en la inútil trampa de enumerar los avatares que la palabra «conocer» haya podido revestir a lo largo de su historia. Me voy a atener al más inmediato del diccionario. En el *María Moliner* (segunda edición, 1998) la primera acepción de la palabra «conocimiento» es: «Acción de conocer». Siguen: «Efecto de conocer o presencia en la mente de ideas acerca de alguna cosa»; «cosas que se saben de cierta ciencia, arte...»; «facultad de saber lo que es o no es conveniente y de obrar de acuerdo con ese conocimiento»; «prudencia, sensatez...» y así, una lista sabia y considerable.

Pienso que la pregunta que se nos hace, tal y como yo la entiendo, no es tan amplia. Si la he comprendido bien, parecería que se nos preguntase si nuestro «conocimiento» –«acción de conocer»; «presencia en la mente de ideas acerca de alguna cosa»– en nuestro campo de actividades –en mi caso la composición musical– pudiera o debiera tener límites, bien por incapacidad de nuestros órganos cognoscitivos –y sus ayudantes diseñados por el hombre– bien por los riesgos que ese «conocimiento» supusiera en caso de uso irresponsable o dañino.

A bote pronto, mi respuesta no puede ser sino una: la música no supone conocimiento alguno de esa índole. No es cuestión de incapacidad ni de riesgo. Simplemente, su ámbito es otro, tan humano como el científico –quizá hasta más necesario para el equilibrio interno del hombre–, pero que responde a otras necesidades o, si se prefiere, cumple otras funciones.

Para ser mejor comprendido, séame permitido formular la pregunta desde el punto de vista de un compositor: algo así como «fronteras de la expresión artística». Sería interesante sometérsela a mis admirados compañeros. Estoy seguro de que yo, al menos, iba a aprender mucho.

Antes de proseguir quisiera aclarar algo. Hay otras «fronteras» a las que no haré sino citar, para que nadie se llame a engaño, pero de las que no hablaré. La música se ocupa de ordenar –o desordenar– sonidos y silencios en el tiempo. Ya sabemos que hay opiniones divergentes. No entraré en ese debate. Está también la «frontera» del sentido de la música en nuestra actual sociedad. Pues bien, la música no es una mercancía, aunque se la trate como tal y compositores e intérpretes aspiremos a vivir de ella. Pero hay músicas nacidas *para ser* una mercancía, incluso –y peor aún– hay excelentes músicas cuyo uso las ha convertido en mercancía. Tampoco entraré en esa selva atornadora –Pascal Quignard habla justamente de *La haine de la Musique* («El odio de la música»)–. Referirse a «fronteras del conocimiento» en estos supuestos es hablar de sociología, no de música. Yo no soy sociólogo. Retomo el hilo.

Si hay algo que quede claro después de consultar cualquier definición de «conocimiento» es que éste se «localiza» —una manera de hablar...— en la conciencia, esto es en el mundo supuestamente racional y consciente: un conocimiento «inconsciente» puede parecer a muchos un juego de palabras, una contradicción, casi un oxímoron.

Sin embargo, para lograr ciertas prácticas que suponen adquirir conocimiento la cosa empieza a no estar tan clara. En el aprendizaje de una lengua, pongo por caso, hay por lo menos dos fases bien diferenciadas. La primera supone un esfuerzo consciente, constante y deliberado de memorización. La segunda podría calificarse de asimilación, en donde lo aprendido con el esfuerzo consciente se transfiere al inconsciente del individuo —los beneficios que acarrea la práctica...—. La lengua ya no requiere deliberación alguna: sale, mejor o peor, con la espontaneidad de lo supuestamente conocido y asimilado.

La música, como cualquier lenguaje, se aprende así en su fase más elemental. Incluso se puede afirmar sin exageración que TODO se aprende así, cuando lo aprendido no va más allá de ese nivel de conocimiento: el práctico.

En esa fase la música es un oficio como cualquier otro —empleo la palabra «oficio» en su sentido más noble, porque lo tiene—. Y como tal su «conocimiento» tiene las fronteras de la eficacia, que van desde lo suficiente hasta lo prodigioso.

Pero tras el «oficio», la música reserva muchas sorpresas, que suponen otros conocimientos, no tanto que aprender, sino que inventar. El compositor formado —y aun antes de formarse— se encuentra frente a la puerta del enigma que le llevó a elegir esa profesión —o sea, cara a cara con su vocación—. Ha aprendido muchas cosas; algunas las juzga inútiles, otras ha preferido olvidarlas —o eso cree—, otras las conserva por si acaso... Pero ahora se trata de encontrar su voz, lo que por definición requiere «otro» oficio, aún por definir, y esa definición —como en «Ante la ley», el breve relato kafkiano— es absolutamente personal: sólo él puede realizarla. Si antes había que asimilar enseñanzas preestablecidas sin rechistar, ahora hay que ser juez y parte: buscar, encontrar, usar y juzgar. El eventual «conocimiento» que tal trabajo puede ofrecerle no será el de un código de reglas heredadas, sino el que se derive de sus propuestas expresivas que, al correr de los años y si están logradas, se acabarán convirtiendo en patrimonio de una colectividad más o menos significativa: reflejo profundo, aunque parcial —hasta hoy no ha habido ninguna capaz de englobar a la Humanidad entera—, pero siempre fiel, de lo que fue ser hombre en un tiempo y lugar determinados (volveré sobre ello).

Con lo que llevo dicho creo que se puede aceptar la tesis de que la música no se ocupa de transmitir «conocimientos» firmes, invariables, ni siquiera evolutivos, sobre nada. En esta situación está acompañada por las restantes artes y sospecho que hasta por las llamadas «ciencias del hombre», aunque éstas lo hagan de distinta manera y por otras razones. En música —en arte en general—, la expe-

riencia que produce ese «conocimiento más allá del oficio» es emocional, incluso para el estudioso; no se diga para el compositor, el intérprete, el público. Una expresión francesa, que por decencia no traduzco y que fue «muletilla» de Stravinski, *ça ne me fait pas bander*, es crudamente gráfica.

Un conocimiento basado en la emoción es siempre sospechoso para un científico —aunque éste haya podido emocionarse al recibir el suyo— y es necesario que así sea. Y ese conocimiento científico es transmisible en palabras o fórmulas. ¿Agota la palabra la posibilidad de comunicarnos, de transmitir? Al margen de las reflexiones sobre «palabra y cosa», «palabra y forma de conocer» etc., en las que no entro por no ser de mi incumbencia, sólo me atrevo a decir que la verbalización excesiva destruye áreas enteras de expresión humana, y que sin duda la música es la demostración elocuente de ello.

Antes insinué que la creación artística —sin duda también la música, aunque apenas si queden restos— ha precedido inconmensurablemente a la ciencia. Con esta reflexión perogrullesca lo que quiero decir es que la forma de «conocimiento» precientífico ha sido el único conocido por el ser humano durante periodos difíciles de medir. Ciertamente, esto no es una justificación. Pero yo no intento justificar nada, sino simplemente señalar realidades de nuestra naturaleza humana que no sería lícito olvidar y, aún menos, combatir. Aquí sí que hay «fronteras», pero no son las de un músico.

El tipo de «sabiduría» —término que quizá sea más justo que el de «conocimiento» para tratar de arte— que la música nos ofrece, se mueve a mi entender entre lo consciente y lo inconsciente: un vaivén constante en el que ambas esferas se enriquecen. Y quizá sea éste uno de sus mayores encantos, incluso la gran razón de su absoluta necesidad.

Aceptemos, siquiera sea como hipótesis, la posibilidad de una «sabiduría» (¿un conocimiento?) de esas características. Aceptemos —y esto sí que es mucho aceptar— que se pueda hablar, escribir, sobre él.

Las preguntas se aglomeran, se multiplican como conejos. Algunas: ¿cómo hablar sobre un arte que no se sirve de las palabras, salvo cuando se trata de técnica u oficio? ¿Cómo calibrar, medir la emoción que produce? ¿Es concebible un repertorio de medios expresivos que provoquen a voluntad ciertos estados de ánimo? ¿Se puede pretender una catalogación de esos medios, caso de que existan? ¿Qué hacer con las incontables formas de expresión que culturas, épocas, han ido acumulando? ¿Son éstas intercambiables o, al contrario, mutuamente incomprensibles? ¿Cómo se puede convertir en colectivo un lenguaje musical? ¿Sería deseable tal cosa? ¿Qué pensar de la vieja monserga de que «la música es un lenguaje universal»? Y así en una inacabable letanía...

Impensable pretender una contestación detallada. E intentar una respuesta globalizadora equivaldría a salirse por la tangente o, simplemente, mentir.

En lo que a mí respecta no puedo sino apuntar algún comentario y, si Erato —o Euterpe, como prefieran, ¡no

Melpómene, por favor!-, si Erato me inspira, digo, opinar con prudencia y discreción sobre ese confuso laberinto.

La pretensión de definir un repertorio de medios musicales capaces de provocar estados anímicos precisos es tan antigua como los primeros documentos de que se dispone. Lo que hace suponer que, en realidad, es tan antigua como la música misma, con o sin documentos. Quizá no sea inútil dar un vistazo a algunos capítulos de su historia —en particular a los más antiguos— desde el punto de vista del material empleado. Esta excursión nos mostrará las sucesivas «fronteras» que el «conocimiento» musical ha experimentado. Confieso tener reparo en presentar este «sobrevuelo» de la música —nada menos—, vista de una manera quizá demasiado personal. Por otra parte, mucho de lo que diga es de dominio público —o eso creo...—. Pido mil perdones por ello, pero no veo otro camino para hablar claramente y con cierto sentido —utilidad, quiero decir— sobre el peliagudo tema que intento tratar.

Los primeros documentos —hindúes, griegos, chinos, tibetanos... un gran etc.— abundan. Casi todos aspiran a lo mismo: bien despertar estados emocionales en el oyente, bien servir de oración. Los medios musicales empleados son variadísimos.

Los griegos estuvieron convencidos del valor expresivo y, sobre todo, ético de la música: el valor, la cobardía, la fidelidad, la molicie etc., podían ser provocados, aumentados, disminuidos por ella —baste con leer a los clásicos—. Hay textos técnicos abundantes. Quizá el más amplio sea el de Aristóxeno de Tarento (siglo IV aC.). A él me atenderé. La idea musical central era la «monodía». Las notas que la componían estaban basadas en la resonancia natural de los cuerpos: intervalos naturales de quinta y su inversión. La monodía se acompañó con la octava y, a partir del siglo IV también de cuartas y quintas, lo que se corresponde con su concepto de «gama» o «escala», organizada en tetracordios, o sea, cuatro notas conjuntas, de las que los extremos son «notas fijas» y las intermedias varían según el tetracordio. De éstos hay tres géneros: diatónico, cromático y enarmónico. No hay diapasón —esto es, alturas absolutas, cosa exclusivamente occidental que llega muchísimo más tarde—. La palabra griega *diapason* designa dos tetracordios sucesivos —o sea una octava—. Las notas que componen ese doble tetracordio llevan ya su nombre e indican su posición en la afinación de la lira. Hay siete especies de octava, cada una con su nombre y sus efectos anímicos. Estos nombres son casi idénticos a los modos eclesiásticos del cristianismo, pero no se corresponden. Por ejemplo, el modo *lidio* griego es:



(de Do a Do), mientras que el eclesiástico es:



(de Fa a Fa).

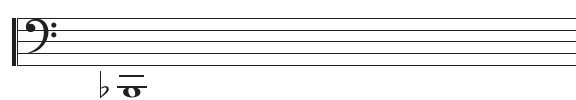
Toda esta teoría —que es detalladísima y que sería enojoso explicar completa— no parece corresponder demasiado a la práctica, que estaba inmersa en una tradición oral sujeta —como todas— a cambios constantes y no demasiado previsibles. Y sin duda esto ha impedido que conozcamos cómo sonaba «de verdad» la música de la antigua Grecia.

En la música profana de la India del Norte —también muy rica en teoría—, los *ragas* —modos, más o menos— y los *talas* —pies rítmicos— son precisos hasta la minuciosidad: hora del día, de la noche; estación del año; estado de espíritu, etc. El material de los *ragas*, o sea, las alturas —intervalos—, se extrae de los 22 grados de que consta su gama completa, y de los que se utilizan casi siempre siete para hacer el *raga* correspondiente. El *raga* necesita un punto de referencia constante para surtir el efecto deseado. Por eso siempre hay un instrumento —la *tampura*, el *ruti*— que funciona como lo que en Europa se llamaría un «pedal» —nota continua—. Naturalmente, los *raga* son numerosísimos: hay *raga* (padre) y *ragini* (madre), que se cruzan y producen innumerable descendencia; *putra* (hijos) y *putri* (hijas)... Los *tala* utilizan dos *tablas* —tambores digitales y manuales—, y están igualmente sometidos a una codificación rigurosa. Y una vez que el aprendiz ha dominado todo este complejo mecanismo, se le da libertad para servirse de él con imaginación —las formas compositivas hindúes estimulan la capacidad improvisativa—. La música de la India del Norte va más allá que la griega en cuanto a efectos: puede curar (o provocar) enfermedades, detener (o provocar) tempestades etc.

La salmodia védica, la música religiosa más arcaica de la India del Norte y quizá el documento musical más antiguo de que se dispone, es posiblemente una de las maneras más intrincadas que puedan imaginarse para contactar con lo divino. La voz del brahmán juega con el texto de mil maneras, todas codificadas: orden de las frases y de las sílabas; acentuación cambiante; sutiles cambios de alturas..., el resultado es incomprensible para el no iniciado. Quizá es eso lo que se pretende: esos cantos de alabanza —que no oraciones— deben ser accesibles sólo a la casta superior.

Por no hacer interminable esta introducción no haré sino aludir a las técnicas vocales del canto religioso tibetano —en este caso sobre escrituras del budismo tántrico— que persiguen el mismo fin, o sea, hacer incomprensible el texto sacro mediante la impostación de la voz masculina en un registro inusual, logrado mediante técnicas muy precisas. Se impide así al profano el acceso a «conocimientos» que se juzgan «peligrosos».

Límite grave de la voz, en torno a:



Sin generalizar, pues, es posible decir que lo conocido de las antiguas culturas musicales muestra su creencia de que la música tiene un poder indudable para provocar estados anímicos variadísimos —que son determinables y determinados— así como capacidad para contactar con lo

divino (ambas cosas pueden ir juntas). Y muestra también que esos poderes y capacidades han revestido innumerables formas y, de paso, han supuesto una primera aproximación para averiguar y utilizar la naturaleza del sonido como fenómeno físico. Y que, detrás de todo ello, brilla una necesidad común de utilizar el sonido como vehículo expresivo, guía u orientación en un universo espiritualizado. Una necesidad tan urgente como la del alimento, la relación sexual, la protección frente a las violencias de la Naturaleza... La música en sus orígenes —tal y como hoy los conocemos— tiene más de misterio que de conocimiento —sería más justo decir que ese conocimiento, catalogado y aprendido rigurosamente, nos permite acceder al misterio de nuestra existencia—. Por eso me he servido de la palabra «sabiduría», que me parece más apta. Pero seguiré utilizando el término «conocimiento» por no apartarme demasiado de nuestro asunto, reservándome el derecho de volver a la «sabiduría» cuando juzgue que ayuda a comprender.

En Occidente, la Iglesia cristiana utilizó los modos griegos de una forma peculiar, a través de lo que llamó *octoechos*: *dorío* o *protus* (re-re), *frigio* o *deuterus* (mi-mi), *lidio* o *tritus* (fa-fa), *mixolidio* o *tetrardus* (sol-sol), con sus correspondientes *plagales* a la cuarta inferior. Los modos se distinguen entre sí por el lugar que ocupa el semitono. En cada modo hay dos notas clave: la *finalis*, para terminar y la *repercussio* (o *corda di recita*) en torno a la cual se construye la melodía. Ésta fue la denominación —errónea respecto de la original griega— que predominó y que luego fue base para la polifonía y el estudio de nuestros armonía y contrapunto. Se excluyeron varios por estimar que algunos de ellos eran demasiado perturbadores para la dignidad requerida en el servicio divino. La música estaba al servicio del texto sacro, y, a diferencia de la India del Norte y el Tíbet etc., éste tenía que ser comprendido por los fieles. Las delicadas inflexiones métricas de los «neumas» debían ayudar al texto y su asimilación. Esta rigidez no impidió la variedad: canto ambrosiano, canto mozárabe, canto galicano, etc., que acabaron uniéndose en el canto gregoriano —muchos tiras y aflojas, nombres propios y riesgos de excomunión.

Hacia el siglo XII se produce en el norte de Francia el nacimiento europeo de la polifonía —cada día gana más adeptos la opinión de que la polifonía nació en África, pero ésa es otra historia—. Desde ese instante la relación entre música y texto empieza a cambiar: éste comienza a ser incomprendible, puesto que se lo confía a más de una voz. Pero la capacidad expresiva de la música aumenta vertiginosamente, con lo cual el monopolio litúrgico de la música se tambalea. Son quizá los trovadores los primeros en hacer una música deliberadamente profana. En el siglo XI ya hay una auténtica «melodía acompañada». Lo encontramos más tarde explícito en Dante («Purgatorio», Canto II): el poeta encuentra a un músico amigo, Casella, que, en vida, había musicado textos suyos. Dante le pide: «... *ti*

piaccia consolare alquanto / l'anima mia, che, con la mia persona / venendo qui, è affanata tanto!». El músico canta el poema de Dante «*Amor che nella mente mi ragiona*», cuya música se ha perdido.

Siento no poder sino aludir al hecho fascinante de que esa independencia de la música respecto de la liturgia —o sea, un texto profano que se plasma en una melodía y se acompaña de algún instrumento— no es europea, sino que vino, por un lado de los cruzados y por otro de la presencia musulmana en España y, en ambos casos, de la antigua Persia —instrumentos incluidos—, de la mano de la disidencia religiosa (los cátaros etc.) con la gloriosa excepción de las «Cantigas», en donde no hubo persecución alguna, quizá por su contenido explícitamente sacro.

El despertar de la música como arte autónomo —esto es, libre de la liturgia e incluso del texto— fue relativamente tardío en Occidente, aunque a través de las sutiles técnicas polifónicas aquella siempre fue considerada como la más emotiva de las artes, imprescindible e inseparable tanto de la poesía como del servicio divino.

Esta interdependencia texto-música religioso-profano no fue la única que nuestro arte conoció. En otras culturas la música corrió muchas aventuras, algunas de las cuales quisiera narrar para mostrar su ductilidad y su lábil —mejor, proteica— naturaleza. Dos ejemplos:

La señora Murasaki (Murasaki Shikibu) nos cuenta en su *Genji* (dinastía Heian, siglo X) cómo había todo un repertorio, a la intemperie, de la flauta travesera *fue* que variaba según las estaciones, teniendo en cuenta el ambiente sonoro natural de cada una de ellas. Esta música se practica aún y yo he tenido la fortuna de escucharla en Kioto, al admirable Suiho Toshia. Obvio es decir que esta idea nace del shintoísmo.

Hay otras músicas que, no contentas con producirse a la intemperie, aspiran a describir minuciosamente algunos aspectos de la realidad: la lluvia, una araña balanceándose en su tela, un niño que llora, un joven que intenta entrar furtivamente en la «casa de las mujeres»... Estoy refiriéndome a la tradición musical de los Aré-aré (Hugo Zemp, «Flûtes de pan mélanésiennes», Musée de l'Homme 1971) y sus conjuntos de flautas de pan (*au tahana* y *au paina*).

He aquí un precioso ejemplo de cómo una música descriptiva, comprensible como tal para sus creadores, no significa absolutamente nada para los ajenos a esa cultura, salvo como objeto exquisitamente bello y desprovisto de significado. Divorcio entre estética y significado originario que encontraremos mil veces.

No quisiera que estas divagaciones —que yo no veo como tales— se tomasen como erudición inútil. Lo que intento con ellas es mostrar la imposibilidad de encontrar «fronteras» a la música que no sean las puramente físicas, de un lado, y de otro que la palabra «conocimiento» no conviene a su naturaleza ni a los efectos que produce en el receptor: su realidad es demasiado varia —no contradictoria— como para pretender una unidad que no haría sino mutilarla.

Bastará con lo dicho sobre este tema.

Cuando se llega en Europa al pleno establecimiento de la música instrumental (más o menos en el siglo xv) las reglas de la composición se precisan con cuidado exquisito, al mismo tiempo que se inventan nuevos géneros en los que la expresión emocional prima sobre lo demás.

A partir de ese momento, se perfila en Occidente algo que sí es únicamente europeo, hasta el extremo de que para muchos aficionados ese «algo» es sinónimo de música. Me refiero —se habrá adivinado— a la armonía, que viene a sustituir a los «modos». La armonía establece la función precisa de los intervalos, midiendo su capacidad de movimiento o de reposo. Su origen es la polifonía que la precede, pero ahora ese movimiento está al servicio de la tonalidad y sus fluctuaciones —las «modulaciones»—. Sin duda la armonía —que ha sido comparada con la perspectiva, ésta en el espacio, aquélla en el tiempo— es la técnica musical más específicamente europea que Occidente ha creado, aunque haya sido a costa de dejarse en el tintero buena parte de la música popular, basada frecuentemente en los viejos modos.

Cuando la armonía se amolda a las reglas del antiguo contrapunto —y viceversa— quedan formulados unos procedimientos que iban a convertirse en la médula misma de la enseñanza europea de la composición. Uno de los hitos de esta enseñanza lo encarna el *Gradus ad Parnasum* (1725), de Johann Joseph Fux. Prácticamente todos los compositores que vinieron después —y durante bastante tiempo— fueron respetuosos con el viejo maestro, aunque se permitieran libertades. Incluso, hace no tantos años, György Ligeti (1978) afirmaba a Péter Várnai (*Ligeti in conversation*, Eulenburg, Londres, 1983) que pedía a sus alumnos de perfeccionamiento en la composición que estuviesen «familiarizados» con el venerable texto, para luego hacer lo que les viniera en gana. Éste es un consejo, dicho sea de paso, que recorre casi todo el siglo xix y, ya vemos, buena parte del xx. Se podría preguntar el porqué, y la respuesta entra en el capítulo del «oficio», al que tantas veces me he referido. Con el Fux en la mano la solidez del edificio está asegurada. Lo que no lo está es el interés que ese edificio pueda tener. Pero eso a Fux, ni a ningún teórico, le concierne: es cosa del artista... que con él ha aprendido «disciplina», aunque sea un indisciplinado, como lo fue, por ejemplo, el joven Debussy. Para decirlo de una vez: el Fux —y sus innumerables parientes— fueron, y quizá aún sean, imprescindibles como polo negativo.

A no tantos años de ser publicado —cerca de cien—, empezaron los embates: el último Beethoven, el romanticismo, la ópera wagneriana, las sutilezas armónicas y tímbricas francesas, los nacionalismos —en particular el ruso— los primeros atonalismos centroeuropeos (no lo que a éstos siguió: ya hablaremos de ello); amén —ya desde otro punto de vista— de las exposiciones universales, que mostraron el arte de las colonias a la metrópoli, etc. Todo ello produjo un fermento imparable que no lo elimi-

nó, pero que cambió su sentido: de norma imprescindible pasó a ser punto de referencia de nuestra vieja identidad, que es bueno conocer si no queremos caminar en falso: un «conocimiento— práctico, no otra cosa (y no es poco).

En lo que acabo de decir se transparenta ya la historia reciente —más o menos los cien últimos años— de la música. El edificio de la armonía clásica —y sus consecuencias formales— se desmorona. Y la causa es tanto interna —su dinámica de desarrollo— como externa —agentes venidos de fuera.

Dos palabras sobre ambos. La causa interna nació en el mismo ámbito cultural del modelo: el mundo germánico. Naturalmente me refiero a la Escuela de Viena y su «trinidad»: Arnold Schonberg, Anton Webern, Alban Berg. Su primer objetivo fue disolver la tonalidad en lo que Schonberg llamó «pantonalidad» —se la conoce también como «atonalidad», término inexacto según él—. Desde 1906-1909, con su *Kammersymphonie* op. 9 —incluso desde su Segundo Cuarteto, op. 7— hasta 1921, en donde define su sistema de composición —que se llamó «dodecafonismo» (o «dodecafonía») en los países latinos; la palabra fue acuñada por el musicólogo franco-lituano René Leibowitz, alumno de Schonberg, y traduce, mejor o peor, el original *zwölftontechnik* o *zwölftonmusik* («técnica» o «música» de doce sonidos)—, los tres vieneses aportan una ingente cantidad de novedades técnicas. Pero Schonberg era además un extraordinario pedagogo, con una formación clásica a toda prueba —pese a ser prácticamente autodidacta—. Su invención del dodecafonismo es una forma de prolongar las técnicas clásicas en un contexto nuevo. Es ejemplar ver ya en su *Pierrot lunaire* (1912), que precede al sistema, cómo se maridan las novedades más arriesgadas con las técnicas más añejas. La Escuela de Viena tenía un agudo sentido de «misión», misión doble y complementaria: de un lado, hacer avanzar con medios nuevos la expresión musical; de otro, recuperar —vivificar— los procedimientos clásicos de la gran escuela germana con contenidos actuales. Así vemos que en el dodecafonismo —aun antes— se retoman el contrapunto severo, las formas clásicas, etc., con el fin de no romper con su pasado, prolongando la supremacía —según ellos— de la música germano-austriaca sobre todas las demás. Pese al nazismo —que prohibió sus músicas como «degeneradas», «bolcheviques», etc., y que obligó al exilio a Schonberg— ninguno de ellos renunció a la idea de considerarse como únicos depositarios de la herencia musical centro-europea. La actitud de la Escuela de Viena fue de una altísima calidad ética, teñida de un nacionalismo un tanto ingenuo, evidentemente no folklórico, pero absolutamente autoexigente. Muy beethovenianamente, su música era su moral —intransigente— frente a un mundo hostil. Las reflexiones agresivas de Karl Popper sobre la música de la Escuela de Viena son un modelo de incompreensión y arrogancia, triste es decirlo.

Entre los agentes foráneos ninguno tan potente como el conocimiento de otras culturas y el cambio de actitud

frente a ellas. Los ataques despectivos de Berlioz, de Gianbatista Vico y de tantos otros, dejan paso a la curiosidad, primero, y al entusiasmo después. Quizá sea difícil encontrar a un compositor europeo de cierta importancia de los últimos setenta años —no los germanoparlantes: ya se ha mencionado el porqué; la excepción es Stockhausen— que no haya sido influido por músicas ajenas a nuestra enseñanza tradicional de la música. Es imperioso reflexionar sobre este hecho.

1. Es Occidente quien ha absorbido estas músicas. Sin duda ha habido países —Japón, Corea, Marruecos, China, etc.— que han asimilado ciertos aspectos de la música occidental. Pero los compositores de esos países lo han hecho desde nuestra tradición. Me explico: Toru Takemitsu, por ejemplo, se sirve de instrumentos, escalas, incluso «sentido temporal» japoneses. Pero el resultado de su obra se inserta en la tradición sinfónica occidental. Su música no se incorpora a la tradición japonesa, no forma parte del *gagaku*, del *kabuki*, del repertorio de los monjes mendicantes con su *shakuhachi* —flauta de pico que él emplea en sus obras—. Séame permitida una anécdota ilustrativa: cuando Toru Takemitsu comenzó a estudiar música en Tokio, su actitud era de rechazo total a la tradición japonesa —los años de la posguerra—. Para ampliar estudios se desplazó a París. Y fue en París, años 60, en donde descubrió la tradición musical de su país y se operó su conversión —esta anécdota es vivida—. Algo muy parecido se puede contar de Ahmed Essyad y Marruecos... y otro tanto de los grandes compositores actuales de Corea, China, etc.

2. Consideradas una por una, estas músicas son intraducibles entre sí. Un ejemplo: difícil es imaginar que la polifonía vocal de los pigmeos *aka* (África Central) pueda enriquecerse con la escucha y estudio de los tambores que acompañan la acción del teatro *Kathakali* de Kerala (Sur de la India), o viceversa. Como digo, estas tradiciones son irreductibles.

Pues bien, esto no reza para un músico occidental. Lo vemos y lo oímos todos los días —teatro y música—. Y la razón es evidente: los occidentales hemos prescindido del contenido extramusical que estas músicas —como todas— conllevan. ¿Qué tiene de raro, si lo hemos hecho con las nuestras? Piénsese en el uso, cada día más frecuente, de técnicas medievales en la música actual. Más aún, piénsese en el neoclasicismo de los años 20 —y en la inmensa figura de Stravinski— y se verá que la cosa viene de antiguo. Y no se vea en lo que digo ninguna crítica negativa, sino un intento de comprender el sentido de una evolución.

3. Nuestra «música de consumo», disponiendo de medios de difusión excepcionales —y convertida por ello en cifra de nuestro bienestar y poder— está invadiendo el planeta y será probablemente lo único que todos los pueblos acaben conociendo de nuestra historia musical. Más aún: la irreductibilidad entre músicas que he señalado, se produce también en Occidente entre nuestra llamada «música culta» o «clásica» —espantosas denominaciones— y nuestra «música de consumo». Basta con saber lo que buena par-

te de la juventud occidental considera como «su» música, «su» vehículo de expresión. Y permítaseme ahora no entrar en consideraciones sobre la procedencia, posibles consecuencias y características musicales de este hecho (al principio de este trabajo pedí permiso para no hacerlo).

En el arranque de este encuentro positivo de nuestra tradición musical con las músicas procedentes de otras culturas —o sea, a principios del pasado siglo— se origina algo, entre nosotros sólo conocido —o casi— por los especialistas, pero que, a mi juicio, tiene gran interés por concernir frontalmente a la interpretación del hecho musical y sus fronteras.

En Alemania y Francia se fundaron centros de estudios musicales que, a la vista de la posibilidad ofrecida por las técnicas de grabación —entonces rudimentarias pero ya eficaces— y de viajar a lugares remotos con cierta seguridad, deciden estudiar la música como fenómeno global y, si es posible, sacar alguna conclusión. En Berlín (1919) Curt Sachs funda el Instituto de Musicología Comparada. En París, André Schaeffner crea el Departamento de Etnomusicología en el Museo del Hombre (1929). En Barcelona, el musicólogo exiliado Marius Schneider hace lo propio en 1944, fundando después (1955) la llamada Escuela de Colonia con el mismo propósito. En mi opinión son estos tres nombres los que definen mejor esa búsqueda, que consiste en esforzarse por encontrar una «parte general» de la música, acopiando una enorme cantidad de datos —en años sin ordenador...— no sólo como hechos aislados, sino incluyendo su contexto y buscando nexos comunes y oposiciones significativas entre materiales dispares. Algunos ejemplos: ¿qué significa el distinto empleo del ámbito vocal en las distintas épocas y culturas? ¿Cómo se ha utilizado el valor cadencial de la caída de cuarta?; comparación entre isocronías y heterocronías, etc. Algunos fueron más lejos. Marius Schneider intenta establecer correspondencias entre ritmos, intervalos, cadencias etc., y signos del Zodíaco, animales, constelaciones, iconografías... André Schaeffner hace lo mismo partiendo de los instrumentos musicales: orígenes, evolución, símbolos, afinación, sentido social —religioso y otro—, etc.

Algunos, en el entusiasmo del descubrimiento de esas analogías/oposiciones llenas de sorpresas, han llegado a pensar que la historia de la cultura se podría conocer e interpretar mejor a través de la música que del lenguaje. En conversación con Alain Daniélou —director del Instituto de Musicología Comparada, años 70— me afirmó rotundamente que las estructuras básicas de la música se habían conservado mejor que las de las lenguas, y que bastaba con la práctica para percibir las. No me siento capaz de opinar sobre este asunto, ciertamente fascinante aunque enigmático. Sí puedo afirmar que, cuando se conocen algunos textos de estos grandes especialistas, la duda asalta al lector de si se trata de un análisis de creencias de un determinado grupo o si estamos ante un intento de resurrección de esas creencias a escala universal,

creencias olvidadas por la negligencia —algunos hablan de «camino equivocado»— de Europa y conservadas en lugares recónditos. Sorprendentemente, algunos musicólogos van ideológicamente mano con mano con Solzhenitsin o monseñor Lefèbvre.

En todo caso este grupo de músicos ha realizado una labor meritísima de conservación, estudio y difusión de músicas en trance de olvido y se ha esforzado por crear una «ciencia de la música» que aspira a dar un sentido global a nuestra historia musical. No han estado solos en estas aspiraciones. Algún fervoroso *sufi* ha pensado lo mismo desde su tradición. Lo que no ha impedido que Karlheinz Stockhausen —y con él un buen número de cultivadores del *rock*— haya estudiado devotamente los escritos de Hazrat Inayat Khan —nacido en lo que hoy es Pakistán a principios del pasado siglo—. Su idea de fondo es la misma: la música es el único vehículo de la sabiduría. Pero su punto de arranque no lo es: los europeos se esfuerzan por encontrar una base científica —*sui generis*, quizá—. No así el asiático, que sólo exige una fe religiosa común. Y más vale que no entremos en el mundo pitagórico, de un lado, y de la antigua China, de otro...

Los procedimientos y resultados de este grupo admirable de musicólogos han sido frecuentemente motejados de «esotéricos», incluso «mágicos», hasta *dilettantes*. Pienso que esta descalificación proviene de que sus trabajos, por interpretar el fenómeno de la cultura a través de la música, han desconcertado a muchos, por lo insólitos. Pero, como músico que soy, a mí no me parecen ni más ni menos «mágicos» que puedan serlo Roman Jakobson con la lingüística o el mismísimo Claude Lévi-Strauss con las estructuras del parentesco. ¿Cuestiones de afinidad? No sólo: cualquier teoría con pretensiones de holística peca de olvidos, cayendo incluso en el ridículo. «El que esté libre de pecado...».

La explosión de la ciencia en el siglo XIX tuvo consecuencias importantes para la música, su sentido y, sobre todo, su conocimiento material —el sonido—. Por ejemplo —no único— el aporte del físico alemán Hermann von Helmholtz (muerto en 1894) fue de gran importancia porque estimuló la imaginación de muchos compositores más de 50 años después de sus descubrimientos. Al analizar la escucha y estudiar los timbres y las alturas, avanzó la hipótesis de que ambos estuvieran relacionados y que podía imaginarse una música en la que esa relación tuviera una función. Por otra parte, preparó las bases de la psicoacústica, lo que marcó un hito en el comportamiento del compositor respecto del material que emplea. No hace falta mucha penetración para percibir que uno de los orígenes de la música de Stockhausen, de la llamada Escuela Espectral francesa y de sus incontables consecuencias, se encuentra en los trabajos de Helmholtz y sus seguidores (muchos sin saberlo).

En paralelo a estas investigaciones científicas no se puede olvidar que los músicos por sí mismos necesitaron enriquecer sus medios expresivos. La sola enumeración de

los ingenios ideados y los resultados logrados ocuparía un espacio abusivo. Citaré los más relevantes. El primero lo constituyen las mil transformaciones de la plantilla instrumental, con énfasis en la percusión —que produjo un sinfín de obras maestras que sería ocioso enumerar y que ya forman parte de la práctica compositiva usual.

No se puede pasar por alto a los *intonarumori* de los futuristas italianos (Luigi Russolo, 1913), que no pasan de ser una curiosidad, pero que indican una postura renovadora.

Con la llegada de la tecnología se produce la auténtica revolución. Primero París, el G.R.M. (Grupo de Investigación Musical con Pierre Schaeffer (1910-1995) y Pierre Henry (1922): la *musique concrète*, en la segunda mitad de los cuarenta; un «arte» que inicialmente se veía a sí mismo como «otro», a mitad de camino entre el cine, el reportaje radiofónico y la música. Acabó decantándose por ésta. Sus materiales de base eran los sonidos/ruidos reales, grabados y trabajados a través de aparatos electroacústicos: magnetófonos, filtros, etc. —incluso se diseñó uno: el *phonogène*, que hoy es una curiosidad de museo, pero que en su día cumplió una función clave—. Poco después, en 1951, Herbert Eimert (1897-1972) y Karlheinz Stockhausen (1928-2007) fundan en la Radio de Colonia el Estudio de Música Electrónica, fabricada a base de generadores de frecuencia, filtros y un largo etcétera.

Ambas, *musique concrète* y *elektronische Musik*, se funden en «música electrónica» o «electroacústica», que se extiende como el consabido «reguero de pólvora» por todo el mundo técnicamente desarrollado: de Estocolmo a Milán, de Lisboa a Varsovia, de Montreal a Buenos Aires, de Tokio a Nueva York, de Sidney a Johannesburgo... Su perfeccionamiento técnico es fulminante. Lo que había nacido como una artesanía se convierte en un mundo de hallazgos en constante evolución. Pronto nacen obras maestras. La más conocida —con justicia— es el *Gesang der Jünglinge* (Cántico de los Adolescentes), de Karlheinz Stockhausen (1955, Colonia).

En 1960 Robert Moog, ingeniero de sonido, diseña en Buffalo (Nueva York) el primer sintetizador Moog, aparato dúctil que permite su uso «en vivo» —esto es, como un instrumento cualquiera—, y es capaz de conexiones múltiples, entre ellas con un ordenador. Su manejo es tan simple que pone al alcance de un público ilimitado los medios electroacústicos. De golpe, lo que había sido considerado como «la punta de la vanguardia» se vulgariza, cayendo directamente en la «música de consumo» —sin abandonar la posibilidad de la creación libre, pero haciéndola progresivamente más difícil.

En 1955, Lejaren Hiller (Nueva York, 1922), en la Universidad de Urbana (Illinois, Estados Unidos), logra que un ordenador, el ILLIAC IV, reconstruya estructuras musicales tradicionales: armonía, forma, etc. El resultado es *Illiac Suite* para cuarteto de cuerda. Según propia confesión —privada—, Hiller no buscaba hacer música, sino demostrar capacidades inéditas de la máquina... Pero muy poco

después se diseña el «convertor digital analógico», con lo que la máquina es capaz de producir sonidos. Hacia 1964, John Chowning, de la Universidad de Stanford (California, Estados Unidos), diseña un ordenador con modulación de frecuencia, con lo que la máquina amplía sus posibilidades a cualquier timbre —instrumental o de invención propia—, métrica, número de voces, etc.

Y aquí es prudente que me detenga. Porque lo que ahora viene —y en ello estamos— es la invasión informática, que pone al alcance de quien lo desee, no ya cualquier material sonoro, sino cualquier cosa relacionada con la música: edición, escucha, mezcla, combinatoria, métrica todo lo compleja que se quiera... Una «cosa» que pueda recordar lo que la música es —por su contenido aural— está disponible hoy para cualquiera que maneje la máquina con cierta habilidad. No hablo de «obra», sino de la posibilidad de hacer «algo». Por lo demás, esa posibilidad no parece ayudar demasiado al profano a conseguir lo que, con todos los respetos, se llama «calidad».

Obvio es que la máquina ha hecho posible cosas que parecían un sueño: transformar un sonido instrumental o vocal en el momento mismo de su producción, desdoblándolo; interacción de alturas en diferentes niveles de escucha; resultados inauditos —en el sentido literal de la palabra, porque no había medios para producirlos—, bien totalmente inéditos, bien de origen conocido, pero irreconocibles; transformaciones de cualquier elemento constitutivo del sonido a velocidad variable..., la lista es inagotable. En los años 50 —en el entusiasmo de la «nueva música»— hubo algún compositor ilustre —cuyo nombre he olvidado— que afirmó con rotundidad y fe conmovedoras que «en diez o veinte años ya no se oíría la música del pasado». No ha sido así, sino que, al contrario, la música llamada «clásica», ahora la del planeta entero —esto es, tanto Bach como la liturgia copta», se escucha más que nunca, con lo que la *facultas eligendi* tiene el trabajo asegurado... Pienso que con el uso del ordenador aplicado a la música sucederá otro tanto: la presencia de un medio tan potente no parece que vaya a impedir la existencia y desarrollo de músicas cuyos medios de expresión sean más tradicionales. Como siempre, lo más probable será la mezcla imprevisible.

Tras este «sobrevuelo» de la música, recorrido apresurado y necesariamente parcial pero, creo, suficiente para la tarea que nos ocupa, hay al menos una cosa clara: los límites del conocimiento musical —entendido como aprendizaje— se han expandido tanto que resulta dudoso decidir qué es lo necesario y qué lo accesorio. Hay que añadir algo importante y de lo que se suele hablar poco: las técnicas —en plural— de composición musical han proliferado hasta lo inverosímil en los últimos 40 años —hablo sólo de Occidente—. Se partió de una ilusoria unidad: el *serialismo*, hijo o nieto de la Escuela de Viena, interpretado de forma muy restrictiva, cuyos focos fueron la llamada Escuela de Darmstadt (Alemania) y sus cursos, fundados

por el Dr. Wolfgang Steinecke en 1946. Técnica severa que saltó hecha añicos a finales de los años 50. Desde entonces no es exagerado decir que hay tantas técnicas de composición como compositores significativos. No existe, o sería mejor decir que nadie ha tenido el valor de hacer, un libro que intente ordenar el inmenso, incalculable, aporte nuevo en la forma, la orquestación, la ordenación interválica, el registro como eje del discurso, la cita... ¿A qué seguir? Si digo simplemente «las incontables maneras de hacer música» seré más veraz y más breve. Hay, cierto, algún intento, sin duda meritorio, pero siempre parcial, bien por referirse a un solo compositor, bien porque el autor no ha podido abarcar todo el asunto. Quizá el único que ha llegado a ser libro de texto en algún conservatorio sea el del compositor polaco Bogusław Schäffer (1929): *Nueva música: problemas de técnica y composición contemporánea* (1958). Como era de temer, no cubre ni la mitad de lo que su título reza. (Uno de los primeros intentos del IRCAM parisino (1976) fue el poner al día un tratado de orquestación. Hasta ahora nada se ha logrado).

No entro en el porqué de la explosión de los 60, que podría ser objeto de otro trabajo. Por lo demás, esa década fue pródiga en acontecimientos revulsivos de toda índole.

Esas tempestades han ido amainando. Desde hace más de veinte años las aguas parecen haberse serenado un tanto. No ha sido por ningún *Rappel à l'ordre* —que hubiera sido tan pretencioso y extemporáneo como el de Jean Cocteau—, sino por una saludable «sístole» tras la pantagruélica «diástole».

Dije antes que los compositores occidentales de los últimos años han (hemos) prescindido del contenido extramusical de la música. O sea, que la música no «representa» más que a sí misma. ¿Es eso así? En mi opinión, sí. Y tendría que añadir: así ha sido siempre, sobre todo en los casos más excelsos, esto es, aquellos en los que la música tiene más poder de conmovir.

Por eso, cuando las músicas no europeas empiezan a ser apreciadas, asimiladas por los occidentales, no lo son por sus contenidos no musicales, sino por la belleza e interés de su materia sonora: nuestros oídos estaban ya preparados.

Nuestros compositores lo han sabido siempre —¿cómo no iban a saberlo, o al menos intuirlo?— incluso cuando hacían música descriptiva. Los madrigales de los siglos XVI-XVII, la ópera, son paradigma de lo que digo. La palabra estimula la imaginación del compositor, pero la música que resulta del estímulo no es una traducción: tarea imposible o frívola. El pasado, quizá con buen criterio, no creó polémica alguna: era evidente que un orden sonoro adecuado era expresivo *per se* si está hecho con imaginación, frescura y maestría. (Olvido intencionadamente la intempestiva querrela de *prima la musica, dopo le parole* —o lo contrario—. No afectó a los compositores).

En nuestros días sí ha habido polémica. Hubo compositores para los que la música debía renunciar a cualquier tipo de expresividad, incluso la propia (recuerdo la

indignación, un poco cómica, de Franco Donatoni contra el poder emotivo del *senza mamma* de la *Suor Angelica* pucciniana). Era también la postura adoptada por el joven Boulez en sus *Structures I*, para dos pianos, de 1952 (no, desde luego, de su *Marteau sans maître*, que las siguió). Pero, en mi opinión, incluso en esa línea compositiva, en la que el autor no se interesa en principio por la dimensión expresiva de un orden sonoro, ese orden, si está logrado (originalidad, perfección, profundidad) nos transmitirá un tipo de emoción, quizá no querida, no buscada, pero que es inherente a su materia y, desde luego, intraducible en palabras. Quizá el error de las *Structures I* (Boulez las ha considerado siempre un experimento) se deba a que el sonido está ordenado de una forma solamente numérico-combinatoria, no musical, con lo cual el resultado es ininteligible. De este mismo error cojean ciertas obras de Xenakis, dicho sea con todos los respetos.

Por otra parte, es más que probable que ese florecimiento de técnicas musicales al que he aludido, tenga uno de sus orígenes en el énfasis que los compositores recientes —y no tan recientes— han dado al material musical *puro* como eje principal del impulso creativo (*oh manes* de Debussy...).

Y nos volvemos a encontrar como al principio: ¿qué se conoce con, por la música? ¿Qué se quiere decir con «Fronteras del conocimiento» cuando se habla de música? Por no tener, la música no tiene ni siquiera la coartada de ese cajón de sastre que es su supuesta capacidad de mejorar la ética de sus fieles, al refinar su sensibilidad. En la mente de todos está la siniestra galería de verdugos, asesinos, torturadores, traidores —de ambos sexos—, etc., que fueron melómanos, protectores de poetas, de pintores... Porque ese «cajón de sastre» no lo es de nadie. El arte no es san Gregorio Taumaturgo. A lo más, lo que hace —y es muchísimo— es enriquecer, sí, la sensibilidad, la imaginación de quien se presta al diálogo. Enseña a oír, a ver, a comprender y sentir de múltiples maneras la realidad que nos rodea. Pero todo esto no mejora nuestra ética. Quizá nos hace más agudos, más ingeniosos (a veces para peor...). La ética no se mejora por ese camino, sino por un esfuerzo personal, individual (con muchísima suerte también colectivo, si se tiene la singular fortuna de un gobierno justo o una comunidad pequeña que no busca el poder), hacia una solidaridad sin relación alguna con la estética.

Evidentemente esto que digo supone para algunos la condena a muerte —sin apelación— del arte. ¿De qué nos sirve, si no nos mejora? O sea: no hay más «mejora» que la ética, al parecer de muchos... Si de verdad creyéramos tal cosa habría que cancelar, desde la revolución industrial hasta la mayoría de las religiones.

El arte es necesario porque forma parte del ser humano como tal: como su nariz, su estómago, o, si se prefiere, su curiosidad por saber. El ser humano exuda «arte» como el caracol su cáscara (aunque no se manifieste en todos por igual. Todos hablamos una lengua. No todos tienen don de lenguas). Y esta «exudación artística» es manifiesta-

mente menos dañina que tantas otras: no implica poder sobre nadie. O si lo da, es mínimo: el «poder» (¿?) de ofrecer obras al prójimo, no de dominarlo. Un poder personal, pero compartible y compartido: si no hay participación no hay arte. Una participación benéfica, que, si degenera en maléfica, es por intrusión de un «poder» ajeno a la obra. No hace falta poner ejemplos.

Un músico puede ser un canalla, un asesino, incluso puede practicar la violencia de género (¡!): no hará daño sino a gente muy próxima (lo que no es una absolución, claro). Una religión o una política dictatoriales, fanatizadas —en Europa hay cierta experiencia al respecto— hacen un daño ilimitado. ¡Que no me hablen de mejora ética a través del arte! Sí de aumento de sensibilidad y de interés por cosas menos nocivas que la televisión comercial.

Las «Fronteras del conocimiento (musical)» serán las de la capacidad del estudioso, del creador, de una parte, y de la capacidad receptiva del contemplador, de otra. En este segundo caso la idea del conocimiento y sus fronteras será idéntica a la capacidad de experimentar una emoción estética profunda, la que nos enriquece personalmente, rara vez colectivamente (¡cuidado con las emociones masivas!), pese a las salas apocalípticas en las que tantas veces músicas nacidas para ser escuchadas a lo sumo por diez personas son escuchadas por diez mil...

Resta por ver si esa «emoción» puede ser una forma de conocimiento. Si lo es —yo así lo creo: véase lo dicho al inicio de este trabajo— pertenece a otra especie cognoscitiva: no busca una verdad objetiva, verificable, por la buena razón de que esa verdad no existe en el arte. Yo me atrevería a afirmar —perdón por meterme en «camisa de once varas»— que es un conocimiento que brota de la experiencia vital, no del estudio. Como he dicho, el estudio —y su goce, porque lo tiene— es para los profesionales. La experiencia vivida, disfrutada, educadora de la sensibilidad —sin excluir al estudioso, ¡faltaría más!— es sobre todo para los demás: aquellos a los que el artista ofrece su obra. Y al correr de los años —casi siempre muchos— eso que un músico dio se convierte en parte de la identidad de un colectivo, parte de su «conocimiento»: se reconoce en él: ésa es la «verdad» del arte, de la música. Dependerá luego de la educación de ese colectivo el que esa verdad sea algo valioso o, al contrario, algo mediocre o trivial. Y, por otra parte, en una sociedad sanamente plural, hay siempre muchos colectivos. Más aún, los colectivos unidos en un común reconocimiento no tienen por qué coincidir con fronteras estatales, religiosas, lingüísticas, ideológicas..., les liga una común emoción, una común «sabiduría» de emociones.

Termino, y lo debo hacer con un interrogante. Hoy por hoy la frontera de esos conocimientos (reconocimientos, sabidurías) está en trance de cambios fulminantes, imprevisibles y posiblemente incontrolados. El músico —y sobre todo el compositor— vive en un constante corto-circuito (ya lo insinué). No es cómodo. Pero no hay tiempo para aburrirse. Por eso lo señalo sin temor... pero con inquietud.

biografías

JANET ABBATE es profesora asociada del departamento de Ciencia y Tecnología de la Sociedad en la Universidad Politécnica de Virginia. Es licenciada en Historia y Literatura por Harvard-Radcliffe y doctora en Civilización Norteamericana por la Universidad de Pensilvania. Es autora de *Inventing the Internet* (MIT Press 1999) y de numerosos artículos sobre la historia de Internet. Como investigadora asociada en el Information Infrastructure Project del Kennedy School of Government de la Universidad de Harvard, coeditó (con Brian Kahin) *Standards Policy for Information Infrastructure* (MIT Press 1995). Sus otras publicaciones incluyen «Privatizing the Internet: Competing Visions and Chaotic Events, 1987-1995» (en *Annals of the History of Computing*, de próxima publicación); «Women and Gender in the History of Computing» (en *Annals of the History of Computing*, 2003); «Computer Networks» (en Atsushi Akeru y Frederik Nebeker eds.). *From 0 to 1: An Authoritative History of Modern Computing*, Oxford 2002), y «Government, Business, and the Making of the Internet» (*Business History Review*, Harvard 2001). Sus investigaciones actuales se centran en la identidad profesional de las mujeres programadoras e ingenieras informáticas en Estados Unidos y Gran Bretaña desde la Segunda Guerra Mundial.

SERGIO ALONSO OROZA es catedrático de Meteorología en la Universitat de les Illes Balears y académico de la Real Academia de Ciencias y Artes de Barcelona. Licenciado y doctor en Física por la Universitat de Barcelona, fue su vicerrector desde 2003 a 2006. Desde 2007 es presidente de la Comisión de Ciencias para la Acreditación Nacional de Profesores Titulares de Universidad. Su labor de investigación se desarrolla principalmente en el campo de la Meteorología y Clima del Mediterráneo, con publicaciones en las revistas más prestigiosas de Meteorología, Oceanografía y Clima. Es miembro de diversas sociedades científicas como la Real Sociedad Española de Física; presidente del Grupo Especializado de Física de la Atmósfera y del Océano; vocal de la Asociación Española de Climatología; y miembro de las Royal Meteorological Society, American Meteorological Society y American Geophysical Union. Entre sus otras actividades destacan: redactor y gestor del Programa Nacional del Clima del Plan Nacional de I+D, miembro de la representación española para el Programa Marco de Naciones Unidas sobre el Cambio Climático, miembro del Consejo Nacional del Clima y de su Comisión Permanente, presidente de la sección de Meteorología y Ciencias de la Atmósfera de la Comisión Española de Geodesia y Geofísica, miembro del Plenario y de la Comisión Ejecutiva y revisor del Intergovernmental Panel on Climate Change.

JESÚS AVILA es profesor de Investigación del CSIC en el Centro de Biología Molecular Severo Ochoa

en la Universidad Autónoma de Madrid, del que fue anteriormente su director. Se doctoró en Ciencias Químicas por la Universidad Complutense de Madrid. Es miembro numerario de la Real Academia de las Ciencias Exactas, Físicas y Naturales y de la Academia Europea. Su trabajo está centrado en el estudio de la función de las proteínas microtubulares en determinar la forma neuronal, en la función de la proteína tau en procesos neurodegenerativos y en la búsqueda de procesos de regeneración axonal. Ha sido presidente y es miembro de la Sociedad Española de Bioquímica, Biología Celular y Biología Molecular, evaluador de proyectos para la National Science Foundation (Estados Unidos), Swiss National Foundation, NFR (Suiza) y Welcom Trust Grants (Reino Unido). Cuenta con numerosos premios, entre los que destacan el de la Real Academia de Ciencias Exactas, Físicas y Naturales, Medalla por la Universidad de Helsinki, Premio de la Fundación Carmen y Severo Ochoa, Medalla de la Comunidad de Madrid, Premio de Investigación Biomédica Preclínica Fundación Lilly y, en 2004, el Premio Nacional de Investigación Ramón y Cajal.

ABHIJIT VINAYAK BANERJEE estudió en las universidades de Calcuta, Jawaharlal Nehru y Harvard, donde se doctoró en 1988. En la actualidad es profesor de la Ford Foundation International, profesor de Economía en el Massachusetts Institute of Technology. En 2003 fundó, en colaboración con Esther Duflo y Sendhil Mullainathan, el Abdul Latif Jameel Poverty Action Lab, del que sigue siendo uno de los directores. Ha sido presidente de la oficina para la Investigación del Economic Analysis of Development, investigador asociado del NBER, investigador invitado del CEPR, profesor internacional invitado del Kiel Institute, amigo de la American Academy of Arts and Sciences, la Econometric Society y de las fundaciones Guggenheim y Alfred P. Sloan. Sus áreas de investigación son la teoría económica y la economía del desarrollo. Es autor de dos libros y editor de un tercero, además de numerosos artículos. En 2006 rodó su primer documental, titulado *The Name of the Disease*.

FRANCISCO CALVO SERRALLER es doctor en Filosofía y Letras por la Universidad Complutense de Madrid, de la que es catedrático de Historia del Arte Contemporáneo desde 1989. Es académico Numerario de la Real Academia de Bellas Artes de San Fernando desde 2001. En 1988 creó la Base de Datos de Arte Contemporáneo ARCO-DATA, que dirigió hasta 1994. Fue director del Museo del Prado. A su labor docente e investigadora se une la de crítico de arte en distintos medios de comunicación, entre otros el diario *El País*, donde colabora desde su fundación en 1976. Ha dirigido y participado en multitud de cursos, congresos y seminarios internacionales sobre arte. Ha formado parte del comité científico

de importantes exposiciones y numerosos certámenes internacionales y ha comisariado importantes muestras en los principales museos y centros de arte de Europa y América. Asesor científico de diversas instituciones, miembro fundador y patrono del consejo de la Fundación Amigos del Museo del Prado y vocal del Consejo Rector del IVAM. Sus investigaciones se han centrado en el estudio de las fuentes de la historia del arte moderno y contemporáneo, en historiografía y metodología artística de la época contemporánea y en la historia del arte contemporáneo.

PAUL E. CERUZZI es conservador de Electrónica Aeroespacial e Informática del Smithsonian's National Air and Space Museum (NASM) en Washington. Es licenciado en Estudios Norteamericanos por la Universidad de Yale y doctor por la de Kansas. Antes de incorporarse al NASM fue becario Fulbright en Hamburgo, Alemania, y enseñó Historia de la Tecnología en la Universidad de Clemson, en Carolina del Sur. Es autor y coautor de varios libros sobre historia de la informática y tecnología aeroespacial: *Reckoners: The Prehistory of The Digital Computer* (1983), *Beyond the Limits: Flight Enters the Computer Age* (1989), *Smithsonian Landmarks in the History of Digital Computing* (1994), *A History of Modern Computing* (1998) y *Internet Alley: High Technology in Tysons Corner* (2008). Sus investigaciones y sus tareas de conservador se centran en el uso de ordenadores en misiones espaciales de largo alcance. Ha comisariado o colaborado en varias exposiciones del NASM (entre ellas: «Beyond the Limits: Flight Enters the Computer Age», «The Global Positioning System: A New Constellation», «Space Race», «How Things Fly») y del James McDonnell Space Hangar del centro Steven F. Udvar-Hazy, en el aeropuerto de Washington-Dulles. En la actualidad prepara una nueva exposición sobre exploración espacial, cuya inauguración en el NASM está prevista para 2010.

CARLOS M. DUARTE GUESADA es profesor de Investigación del Consejo Superior de Investigaciones Científicas en el Instituto Mediterráneo de Estudios Avanzados (IMEDEA). Licenciado en Biología Ambiental en la Universidad Autónoma de Madrid, tras dos años de investigación en Portugal completó su tesis doctoral sobre ecología de macrófitos de lagos en la Universidad McGill de Montreal. Después de una breve estancia posdoctoral en la Universidad de Florida, inició su andadura en el estudio de ecosistemas marinos en el Instituto de Ciencias del Mar (CSIC), trabajó en el Centro de Estudios Avanzados de Blanes (CSIC) y, finalmente, en el Instituto Mediterráneo de Estudios Avanzados (CSIC-Univ. Illes Balears). Su investigación se ha extendido a un amplio espectro de ecosistemas acuáticos. Ha publicado más de 300 artículos científicos en revistas internacionales, una docena de capítulos en diver-

sos textos y dos libros. Es presidente de la Sociedad Americana de Limnología y Oceanografía, editor jefe de la revista *Estuaries and Coasts*, miembro del Comité Científico Asesor del CSIC y de la Academia Europea. En 2001 recibió la Medalla G. Evelyn Hutchinson a la excelencia científica por la American Society of Limnology and Oceanography, y en 2007 el Premio Nacional de Investigación Alejandro Malaspina.

JOAN ESTEBAN es investigador del Instituto de Análisis Económico del CSIC desde su fundación en 1985 y ha sido director del mismo entre 1989 y 1991 y entre 2001 y 2006. En la actualidad es presidente de la Asociación para el Estudio de la Economía en Igualdad y miembro del consejo ejecutivo de la *International Economic Association* (2005-2011). Sus trabajos de investigación se han publicado en *American Economic Review*, *American Political Science Review*, *Econometrica*, *Economics of Governance*, *Economics Letters*, *European Economic Review*, *International Economic Review*, *Journal of Economic Behaviour and Organization*, *Journal of Economic Theory*, *Journal of Income Inequality*, *Journal of Peace Research*, *Regional Science and Urban Economics*, *Social Choice and Welfare* y *Theory and Decision*. En el presente sus áreas de investigación son el estudio de los conceptos de conflicto y la polarización, temas que abarcan diversos aspectos, desde el comportamiento individual y de grupo a la función de los mecanismos políticos en la prevención de conflictos abiertos. Otros campos de interés son las finanzas públicas, la desigualdad, los modelos OLG, el comportamiento en situaciones de incertidumbre y la economía regional.

LUIS FERNÁNDEZ-GALIANO es arquitecto, catedrático de Proyectos en la Escuela de Arquitectura de la Universidad Politécnica de Madrid y director de la revista *AV/Arquitectura Viva*. Miembro de número de la Real Academia de Doctores, ha sido Cullinan Professor en la Universidad de Rice, Franke Fellow en la Universidad de Yale, investigador visitante en el Centro Getty de Los Ángeles y crítico visitante en Harvard y Princeton, así como en el Instituto Berlage; y ha dirigido cursos de las universidades Menéndez Pelayo y Complutense. Presidente del jurado en la 9ª Bienal de Arquitectura de Venecia y de la XV Bienal de Arquitectura de Chile, experto y jurado del Premio europeo Mies van der Rohe, ha sido comisario de las exposiciones «El espacio privado» en Madrid y «Eurasia Extrema» en Tokio, y ha formado parte del jurado de numerosos concursos internacionales. Entre sus libros se cuentan: *La quimera moderna*, *El fuego y la memoria*, *Spain Builds* (en colaboración con el MoMA) y *Atlas, arquitectura global circa 2000* (con la Fundación BBVA).

JOHN B. HEYWOOD trabaja desde 1986 en el Massachusetts Institute of Technology, donde es Sun Jae Professor de Ingeniería Mecánica y director del Sloan Automotive Laboratory. Sus intereses se centran en los motores de combustión interna, sus combustibles y el futuro de la tecnología de transportes, los combustibles alternativos, la contaminación atmosférica y la emisión de gases de efecto invernadero. Ha publicado más de 200 artículos especializados y es autor de varios libros, incluido el manual de referencia imprescindible para profesionales titulado *In-*

ternal Combustion Engine Fundamentals. Pertenece a la Society of Automotive Engineers. Ha recibido varios premios, incluidos el US Department of Transportation Award for the Advancement of Motor Vehicle Research and Development en 1996 y el Society of Automotive Engineers 2008 Award por sus contribuciones a las regulaciones de la industria automotriz. Es miembro de la National Academy of Engineering y amigo de la American Academy of Arts and Sciences. Es doctor por el MIT y por la Universidad de Cambridge y doctor honoris causa por la Universidad Chalmers de Tecnología de Suecia y la City University de Londres.

GERALD HOLTON es profesor investigador de Física e Historia de la Ciencia de la Universidad de Harvard, miembro de las American Physical Society, American Philosophical Society, American Academy of Arts and Sciences y de otras sociedades similares europeas. Ha sido presidente de la History of Science Society y de varias comisiones nacionales en Estados Unidos. Entre sus libros publicados en España figuran *Fundamentos de la física moderna* (1972), *Ensayos sobre el pensamiento científico en la época de Einstein* (1982), *Introducción a los conceptos y teorías de las ciencias físicas* (1984) y *Ciencia y anticiencia* (2003). Fue editor fundador de la publicación trimestral *Daedalus* y miembro del comité editorial de *Collected Papers of Albert Einstein*. Entre sus distinciones están la medalla Sarton y el Abraham Pais Prize de la American Physical Society. Asimismo ha sido Herbert Spencer Lecturer de la Universidad de Oxford University, Jefferson Lecturer por el National Endowment for the Humanities y Ehrenkreuz 1.Klasse en Austria.

CAYETANO LÓPEZ cursó sus estudios de Física en las universidades Complutense y París VII. Doctor por la Universidad Autónoma de Madrid, es catedrático de Física Teórica de dicha universidad desde 1983. Es doctor honoris causa por la Universidad de Buenos Aires. Ha trabajado en el CNRS (Centre National de la Recherche Scientifique) y en el CERN (Centre Européen de la Recherche Nucléaire) en campos relacionados con las teorías unificadas de partículas elementales y la transmutación nuclear con fines energéticos y de eliminación de residuos radiactivos. Además de numerosos artículos en revistas especializadas y contribuciones sobre divulgación científica, política científica y energía en diarios y revistas de pensamiento, es autor de dos libros de divulgación, *El ogro rehabilitado* (El País-Aguilar, 1995) y *Universo sin fin* (Taurus, 1999), y un libro de texto, *Física de los procesos biológicos* (Ariel, 2004). Ha sido rector de la Universidad Autónoma de Madrid, vicepresidente del Consejo del CERN y director del Parque Científico de Madrid. Desde septiembre de 2004 es director general adjunto del CIEMAT (Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas), donde dirige el Departamento de Energía.

JOAN MASSAGUÉ se doctoró en Bioquímica y Farmacia por la Universidad de Barcelona en 1978. En 1982 se unió al equipo docente de la Facultad de Medicina de la Universidad de Massachusetts y en 1989 fue nombrado presidente del Memorial Sloan-Kettering Cancer Center. Es investigador del Howard Hughes Medical Institute y director adjunto del Instituto de Investigación Biomédica de Barcelona. Sus trabajos

han revelado cómo las señales externas bloquean la división de células mamarias. En la actualidad se sabe que estos mecanismos con cruciales en el desarrollo embrionario y que su alteración causa trastornos congénitos y cáncer. A partir de la posibilidad de diseccionar procesos biológicos complejos, el doctor Massagué ha identificado recientemente genes y funciones que median en la formación de metástasis, principal agente de mortalidad en enfermos de cáncer. Considerado uno de los cincuenta mejores científicos de las últimas dos décadas, el doctor Massagué es además miembro de la National Academy of Sciences, la Academia Nacional de Ciencias, el Institute of Medicine, la American Academy of Arts and Sciences, las reales academias españolas de Medicina y Farmacia y la European Molecular Biology Organization. Ha recibido, entre otros galardones, el premio Príncipe de Asturias de investigación científica y técnica, el Vilcek Prize y el Passano Prize.

JOSÉ M. MATO es director general de CIC BIOGUNE y CIC biomaGUNE en el País Vasco desde 2003. Realizó el doctorado en la Universidad de Leiden y, antes de establecerse en el País Vasco, fue investigador de la Fundación Jiménez Díaz (Madrid), profesor de investigación del Consejo Superior de Investigaciones Científicas (CSIC), presidente del CSIC y catedrático de la Universidad de Navarra. Ha sido investigador invitado de los National Institutes of Health, de la Universidad de Carolina del Norte en Chapel Hill y de las universidades de Pennsylvania y Thomas Jefferson, en Filadelfia. Ha recibido numerosos premios, entre los que se encuentran el Premio Kok, la Medalla Morgagni, el Premio Lennox K Black y, en 2004, el Premio Nacional de Investigación en Medicina Gregorio Marañón, el mayor reconocimiento científico que concede el Estado español. El trabajo del profesor Mato está centrado en el estudio de la enfermedad del hígado graso, la regeneración hepática y el cáncer de hígado. Su laboratorio ha sido pionero en la identificación de la S-adenosilmetionina como regulador fundamental del metabolismo y la proliferación hepática, así como en la creación de nuevos modelos animales para el estudio de las enfermedades hepáticas.

ROBERT MCGINN es profesor de Ciencias de la Gestión e Ingeniería y director del programa de Ciencia, Tecnología y Sociedad de la Universidad de Stanford. Se licenció en Ciencia y Tecnología Unificadas en el Stevens Institute of Technology, realizó estudios de posgrado de Matemáticas en Stanford y se doctoró en Filosofía e Historia de las Ideas también en Stanford. Aparte de una breve estancia en los laboratorios Bell (1978-1979), ha trabajado en Stanford desde 1971. Sus áreas de especialización académica son tecnología y sociedad; ética, ciencia y tecnología; ética de la ingeniería y ética y legislación. Sus publicaciones incluyen *Science, Technology and Society* (1991), así como artículos en revistas especializadas como *Professional Ethics; Technology and Culture; Science, Technology and Human Values; Science and Engineering Ethics* y *Nanoethics*. Entre 2004 y 2007 dirigió un estudio de gran envergadura para el US National Nanotechnology Infrastructure Network sobre las opiniones de investigadores sobre cuestiones éticas relacionadas con la nanotecnología. Ha sido galardonado con varios premios a la excelen-

cia docente en Stanford, también ha recibido becas de investigación de la Mellon Foundation, el Marshall Fund y el National Endowment for the Humanities. Asimismo, ha sido Premio Dinkelspiel por sus contribuciones a la educación universitaria en Stanford.

GINÉS MORATA es licenciado en Ciencias por la Universidad Complutense de Madrid (1968) y doctor por la misma Universidad en 1973. Durante los años 1973 y hasta 1977 realizó trabajos de Investigación en las universidades inglesas de Oxford y Cambridge. En 1977 se incorporó al Centro de Biología Molecular del Consejo Superior de Investigaciones Científicas y de la Universidad Autónoma de Madrid, donde continúa en la actualidad con el cargo de profesor de Investigación. En el periodo 1990-1991 fue director del citado centro. El profesor Morata es un especialista en Genética del Desarrollo, disciplina científica en la que lleva trabajando más de 30 años. Hasta el presente ha publicado 115 artículos de investigación en revistas científicas internacionales y ha sido invitado a pronunciar conferencias sobre su labor científica en numerosas universidades y centros de Investigación de Europa, América y Asia. Ha dirigido, además, diez tesis doctorales. Ha sido galardonado con varios premios y distinciones, entre los que destacan el Premio Rey Jaime I de Investigación en Genética 1996, el Premio Nacional de Investigación Santiago Ramón y Cajal 2002, el Premio México de Ciencia y Tecnología 2004 y el Premio Príncipe de Asturias de Investigación Científica y Técnica 2007. Es, además, doctor honoris causa por las universidades de Alcalá (2007) y Almería (2008).

LUIS DE PABLO estudió composición con Max Deutsch (París) y recibió cursos en Darmstadt (Alemania). Es doctor honoris causa por la Universidad Complutense. Tiene más 150 obras que han sido interpretadas por el cuarteto Arditti, Pierre Boulez, Bruno Maderna, ONE, la orquesta de París, la Metropolitana de Tokio, Claude Helffer, José Ramón Encinar, Rafael Frühbeck, Massimiliano Damerini, la orquesta SWF Baden-Baden, NDR de Hamburgo, la Filarmónica de Berlín o el Trío Arbós. Es profesor de composición en Buffalo (Nueva York), Ottawa, Montreal, Madrid, Milán y Estrasburgo, académico de las Academias de Bellas Artes de Madrid y Granada, de Santa Cecilia (Roma) y de la Real Academia Belga, oficial de las Artes y Letras de Francia y miembro de la Sociedad Europea de la Cultura desde 1966. En 1964 funda el primer Laboratorio de Música Electrónica de España y en 1965 el centro ALEA. Entre sus numerosos premios cuenta con la Medalla de Oro del Rey, la de la Cruz Roja Española, la del Círculo de Bellas Artes (Madrid), el Premio Fundación Guerrero, el Premio CEOE a las Artes, las medallas de Rennes y Lille (Francia), el Premio Honegger (París) y el Premio Pierre de Mónaco. Hay cursos sobre su obra en París, Madrid, Ueno Gakuen (Tokio), Instituto di Tella (Buenos Aires), Academia de Santa Cecilia (Roma), Elisabeth School of Music (Hiroshima) y UCLA (Los Ángeles); y el INA (Instituto Nacional de l'Audovisuel) le hizo una entrevista para su archivo de imagen y sonido. Su música está publicada en TONOS (Darmstadt), Salabert (París) y Suvini Zerboni (Milán).

NATHAN ROSENBERG es profesor emérito Fairleigh S. Dickinson de Legislación Pública en la

facultad de Economía de la Universidad de Stanford y Senior Fellow del Stanford Institute for Economic Policy Research. Sus investigaciones están centradas en la economía del cambio tecnológico. Sus publicaciones versan en su mayoría sobre las causas y las consecuencias del cambio tecnológico. También ha examinado la diversidad de las fuerzas que generan el cambio tecnológico en distintos sectores de la industria así como las influencias mutuas entre investigación e innovación tecnológica. Sus libros incluyen *The American System of Manufactures, Perspectives on Technology, Progreso técnico: el análisis histórico, Dentro de la caja negra: tecnología y economía y Exploring the Black Box, The Emergence of Economic Ideas and Schumpeter and the Endogeneity of Technology*. Es doctor honoris causa por las universidades de Lund, Bolonia y Northwestern, y premio Leonardo da Vinci por sus contribuciones a la historia de la tecnología.

VICENTE SALAS FUMÁS es Ph.D. en Management por la Universidad de Purdue y licenciado en Administración de Empresas y MBA por ESADE. Actualmente es catedrático de Organización de Empresas en la Universidad de Zaragoza y lo ha sido también en la Universidad Autónoma de Barcelona. Sus líneas de investigación incluyen el análisis económico de las organizaciones y los estudios empíricos sobre las empresas españolas. Ha publicado artículos en revistas académicas nacionales e internacionales y libros como *Economía de la empresa: decisiones y organización, El gobierno de la empresa, La empresa familiar en España y El siglo de la empresa*. En 1992 recibió el Premio Rey Jaime I de Economía.

FRANCISCO SÁNCHEZ MARTÍNEZ es uno de los pioneros y promotores de la Astrofísica en España. Primer catedrático de Astrofísica de la universidad española, funda y dirige el Instituto de Astrofísica de Canarias (IAC), gracias al cual desarrolla y divulga esta ciencia y sus tecnologías conexas en el país. Hace posible la construcción del mayor y más avanzado telescopio del momento: el Gran Telescopio Canarias de espejo segmentado y más de diez metros de diámetro. Pertenece a diversas sociedades científicas internacionales y tiene múltiples distinciones, entre ellas: Encomiendas de Alfonso X el Sabio, del Mérito Civil y de Isabel la Católica, Cruz del Merito Naval, Gran Cruz de la Orden Islas Canarias, Commander of the Royal Order of the Polar Star (Suecia), Commandeur dans l'Ordre des Palmes Académiques (Francia), Caballero de la Orden Real de Orange-Nassau (Países Bajos), Commandatore de la Stella della Solidarietà (Italia); además es doctor honoris causa por las universidades de Copenhague y Florida, Medalla de Honor al Fomento de la Invención de la Fundación García Cabrero, Premio CEOE a las Ciencias, Medalla de Oro de Canarias y Premio Canarias de Investigación.

JOSÉ MANUEL SÁNCHEZ RON es licenciado en Ciencias Físicas por la Universidad Complutense de Madrid y doctor (Ph.D.) en Física por la Universidad de Londres. En la actualidad es catedrático de Historia de la Ciencia en el Departamento de Física Teórica de la Universidad Autónoma de Madrid, en la que antes fue profesor titular de Física Teórica. Desde

2003 es miembro de la Real Academia Española, en la que ocupa el sillón G. Es, asimismo, miembro de la Academia Europea de Ciencias y Artes (Academia Scientiarum et Artium Europaea) y académico correspondiente de la Real Academia de Ciencias Exactas, Físicas y Naturales y de la Académie Internationale d'Histoire des Sciences (París). Es autor de más de 300 publicaciones en los campos de la física teórica, historia y filosofía de la ciencia, entre las que figuran una treintena de libros como: *El origen y desarrollo de la relatividad* (1983), *Diccionario de la ciencia* (1996), *Cinco, martillo y piedra* (1999), *El siglo de la ciencia* (2000), por el que recibió el Premio José Ortega y Gasset de Ensayo y Humanidades de la Villa de Madrid, *Historia de la física cuántica. I* (2001), *El jardín de Newton* (2001), *El poder de la ciencia* (2007) y *¡Viva la ciencia!* (con Antonio Mingote; 2008).

ANGELIKA SCHNIEKE y **ALEXANDER KIND** trabajan en colaboración desde 1984. Angelika Schnieke trabajó con Rudolf Jaenisch entre 1978 y 1987, primero en el Instituto Heinrich-Pette de Hamburgo y después en el Massachusetts Institute of Technology, investigando las transgénesis retroviral en ratones. Alexander Kind se doctoró en 1984 en la Universidad de Cambridge, donde trabajó con Sir Martin Evans en la diferenciación temprana de células madre embrionarias. Más tarde se unió al equipo de Rudolf Jaenisch para trabajar en el desarrollo de modelos animales para trastornos del tejido conectivo. Schnieke y Kind han extendido sus investigaciones a la producción de ganado transgénico en la Universidad Estatal de Colorado. Entre 1992 y 2003 trabajaron con la compañía biotecnológica PPL Therapeutics en Edimburgo, donde Angelika Schnieke obtuvo su título de doctora por su tesis sobre el desarrollo de transferencias de células nucleares somáticas como medio de generar ovejas transgénicas. En la actualidad Schnieke es catedrática de Biotecnología del Ganado en la Technische Universität de Múnich. Su investigación y la de Kind se centra en este momento en células madre animales y en la ingeniería genética del ganado para su aplicación biomédica.

SANDIP TIWARI nació en la India y se educó en IIT Kanpur, RPI y en la Universidad de Cornell. Después de trabajar como investigador en IBM se unió al equipo docente de Cornell en 1999. Ha sido profesor visitante en las universidades de Michigan, Columbia y Harvard, y fundador y editor jefe de *Transactions on Nanotechnology*. Es autor de un manual de referencia sobre física de dispositivos. En la actualidad es profesor Charles N. Mellowes de Ingeniería y director del National Nanotechnology Infrastructure Network de Estados Unidos. Sus áreas de investigación abarcan la ingeniería y la ciencia de dispositivos ópticos y semiconductores. Ha recibido los premios Cleo Brunetti de la Institution of Electronic and Electrical Engineers (IEEE), el Distinguished Alumnus Award del IIT Kanpur, el Young Scientist Award del Institute of Physics, así como becas de investigación de la American Physical Society y el IEEE. En la actualidad investiga sobre los desafíos planteados cuando se conectan escalas de gran tamaño, como las de los sistemas electrónicos integrados, con escalas reducidas, como las de dispositivos de pequeño tamaño. También estudia las estructuras que resultan del uso de la nanoescala.

créditos imágenes

Cubierta: Nasa, Esa y T.M. Brown (STScI). Estrellas jóvenes y viejas encontradas en el halo de Andrómeda.

p. 6: Nasa Solarsystem Collection. Imagen en 3D del Sol tomada en la onda ultravioleta por la sonda Stereo.

p. 10: Candida Höfer/VG Bild-Kunst, Bonn, 2008. *Biblioteca Geral da Universidade de Coimbra III*, 2006, 260 x 200 cm.

pp. 14–15: Theo Allofs/Corbis. Glaciar de Perito Moreno.

pp. 20–21: Collart Herve/Corbis Sygma. Escenas atmosféricas al atardecer en la selva del Amazonas, Bazin, Brasil.

pp. 26–27: Nasa/The Hubble Heritage Team (STScI/Aura/Nasa). Nebulosa del Anillo del Sur.

p. 30: Oliver Meckes and Nicole Ottawa/Science Photo Library/AGE. Imagen termográfica de un secador de pelo eléctrico que muestra la distribución del calor en el difusor e ilustra las variaciones de temperatura en la superficie.

pp. 40–41: Harry Gruyaert/Magnum Photos. La BBC II graba el Apolo XII, Londres, 1969.

p. 62: Professor Harold Edgerton/Science Photo Library/AGE. *Milk Drop Coronet* [corona de gota de leche], 1957.

pp. 72–73: Professor Harold Edgerton/Science Photo Library/AGE. *30 Bullet Piercing an Apple* [Bala taladrando una manzana], 1964.

pp. 84, 85, 86, 87, 70, 75, 81, 82, 84, 85 y 87: AIP Emilio Segrè Visual Archives.

p. 92: Oliver Meckes and Nicole Ottawa/SciencePhoto Library/AGE. Microengranajes.

p. 97: Oliver Meckes and Nicole Ottawa/Science Photo Library/AGE. Micro-submarino en una arteria.

pp. 104–105: Image 100/Corbis. Amasijo de cables con un conector transparente.

p. 108: Gustoimages/SciencePhoto/AGE. Rayos X a color de un ordenador portátil.

pp. 118–119: José Azel/Aurora/Asa. Los monitores de ordenador emiten una luz verde brillante.

p. 128: Peter Menzel/Asa. Vista nocturna de una antena tipo plato del observatorio VLA, cerca de Socorro, Nuevo México.

pp. 134–135: Peter Menzel/Asa. Observatorio en Mount Hamilton, San José, California.

p. 142: Jupiter images/Brand X/Corbis. Circuito al espacio exterior.

pp. 148–149: Nina Berman/Aurora Pictures/Asa. Earthlink NOC en Atlanta.

pp. 156–157: Adam Fuss. *Invocation*, 1992. Fotograma único en cibachrome, 101,6 x 76,2 cm. Cortesía de Cheim & Read, Nueva York.

p. 160: Andrew Brooks/Corbis. Franjas de ADN en una pantalla de ordenador, junio de 2001.

p. 166: Peter Menzel/ASA. Micrografía fluorescente de cromosomas humanos.

p. 170: Frans Lanting/Corbis. Vasos sanguíneos de la mano.

p. 177: Roger Ressmeyer/Corbis. Cultivos de fotobacteria NZ-11 brillando en platos de petri, septiembre de 1983.

p. 184: Najlah Feanny/Corbis. Dolly, febrero de 1997.

pp. 190–191: Visuals Unlimited/Corbis. Óvulo y espermatozoide humanos, aumentados al 2.500%.

p. 194: RBM online/epa/Corbis. Primer embrión clonado en Gran Bretaña.

p. 202: Microdiscoveries/Corbis. Célula cancerosa de colon, aumentada al 2.000%.

pp. 208–209: Visuals Unlimited/Corbis. Micrografía confocal de células humanas inmortales cultivadas (HeLa) teñidas para observar la distribución de la proteína b-tubulina (verde) y la f-faloidina (roja) en el citoesqueleto, así como el ADN en el núcleo celular (azul). Microscopía fluorescente de dos fotones aumentada al 1.800%.

pp. 222–223: Ingo Arndt/Minden Pictures/Asa. Patas de animales.

p. 224: Sanna Kannisto. *Heliconias*, 1998, c-print, 105 x 130 cm.

p. 230: Sanna Kannisto. *Rothschildia*, 2001, c-print, 74 x 90 cm.

p. 231: Sanna Kannisto. *Dasyus Novemcinctus*, 2001/2008, c-print, 74 x 94 cm.

p. 234: Lester Leftkowitz/Corbis. Paso elevado en hora punta.

pp. 248–249: Harry Gruyaert/Magnum Photos/Contacto. Industria Renault en Turquía.

p. 256: Peter Marlow/Magnum Photos/Contacto. Estación nuclear en Dungeness, costa de Kent.

p. 263: Ian Berry/Magnum/Contacto. Plataforma costera británica de gas natural, Tailandia.

pp. 272–273: Nasa/Earth Observatory Collection. Imagen de la sonda de Lancaster en el pasaje noroccidental entre Canadá y Groenlandia capturada por el MODIS (Espectrorradiógrafo de Imagen de Resolución Moderada) en el satélite Terra, septiembre de 2007.

p. 276: Gideon Mendel/Corbis. Inundaciones en Bihar, India, agosto de 2007.

pp. 296–297: Pablo Zuleta Zahr. *Hombres de negro*, proyecto *Baquedano*, 2006.

p. 300: Stuart Franklin/Magnum/Contacto. Lloyds, Londres, 1999.

pp. 304–305: Jacqueline Hassink. *Mesa de reunión de la junta directiva de BMW*, Stuttgart, Alemania, 15 de diciembre de 1994. *Mesa de reunión de la junta directiva de Daimler Benz*, Stuttgart, Alemania, 25 de agosto de 1994. *Mesa de reunión de la junta directiva de EDF Electricité de France*, París, Francia, 30 de noviembre de 1994. *Mesa de reunión de la junta directiva de Philips Electronics*, Eindhoven, Países Bajos, 15 de julio de 1993.

p. 308: Gueorgui Pinkhassov/Magnum Photos/Contacto. Oficina.

p. 314: Peter Marlow/Magnum/Contacto. Centro de poder global, la City de Londres, la LME (London Metal Exchange), 1999.

pp. 320–321: Erich Hartman/Magnum Photos/Contacto. Cajas de seguridad, Zúrich.

p. 328: Peter Marlow/Magnum/Contacto. Londres, 1999.

pp. 332–333: Harry Gruyaert/Magnum/Contacto. *La Bolsa*, 1998.

p. 336: Hiruji Kubota/Magnum/Contacto. Sede de la dirección de la Hong Kong and Shanghai Banking Corporation (hoy Hongkong Bank), diseñada por el arquitecto británico sir Norman Foster, 1996.

pp. 342–343: Tyler Hicks/The New York Times/Contacto. Afganistán.

p. 346: Bruno Barbey/Magnum/Contacto. *El río Amazonas*, 1966.

pp. 352–353: Steve Mc Curry/Magnum Photos/Contacto. Hombre empujando un carro delante de una casa destruida en un incendio, Kabul, Afganistán, 1985.

pp. 362–363: Reuters/Carlos Barria/Cordon Press. Un hombre pasa delante de una pintura del Contemporary Fine Arts de Berlín durante la exposición internacional Art Basel en Miami Beach, 2005.

p. 366: Eva-Lotta Jansson/Corbis. Visitantes ante la gigantesca instalación *Sun [Sol]*, de Olafur Eliasson, parte del *Weather Project* en la Tate Modern, Londres, noviembre de 2003.

p. 370: Bruno Barbey/Magnum Photos/Contacto. *Caballo colgado*, Maurizio Cattelan.

p. 374: Jeff Goldberg/ESTO.

p. 376: Roland Halbe.

p. 377: Margherita Spiluttini.

p. 379: Michel Denancé.

p. 381: Michael Moran.

p. 383: Nacása Et Partners.

pp. 384–385: Iwan Baan.

p. 386: Nigel Young/Foster and Partners.

p. 390: Colin McPherson/Corbis. Violines dañados expuestos en el taller de Andrew Hutchinson en Hoylake, noroeste de Inglaterra, abril 2008.

pp. 396–397: Candida Höfer/VG Bild-Kunst, 2008. *Stadtcasino Basel I*, 2002, 152 x 204 cm.

Ninguna parte de esta publicación puede reproducirse o transmitirse de ninguna forma, o por ningún medio electrónico o mecánico, incluidas la fotocopia, la grabación o cualquier otro sistema de almacenaje y recuperación de información, sin el permiso previo por escrito del editor.

Edición
BBVA

Dirección y coordinación del proyecto
Gabinete de Presidencia BBVA
Área de Comunicación e Imagen BBVA
Departamento de Actividades Culturales

Producción editorial
Turner

Coordinación editorial
Cecilia Gandarías

Concepto editorial y textos adicionales
Carlos Wert

Textos
Janet Abbate, Sergio Alonso, Jesús Avila, Abhijit V. Banerjee, Francisco Calvo Serraller, Paul E. Ceruzzi, Carlo M. Duarte, Joan Esteban, Luis Fernández-Galiano, John B. Heywood, Gerard Holton, Alexander Kind, Cayetano López, Joan Massagué, José M. Mato, Robert McGinn, Ginés Morata, Luis de Pablo, Nathan Rosenberg, Vicente Salas Fumás, Francisco Sánchez Martínez, José Manuel Sánchez Ron, Angelika Schnieke y Sandip Tiwari

Traducciones
Laura Vidal y Mar Ozores
Sirk traducciones (texto clonación)

Edición textos
Ana Muncharaz

Edición gráfica
Lala Herrero

Diseño
Estudio Joaquín Gallego

Maquetación
Pablo Nanclares
Concha Pérez
Nicolás García Marque

Fotomecánica e impresión
Brizzolis, arte en gráficas

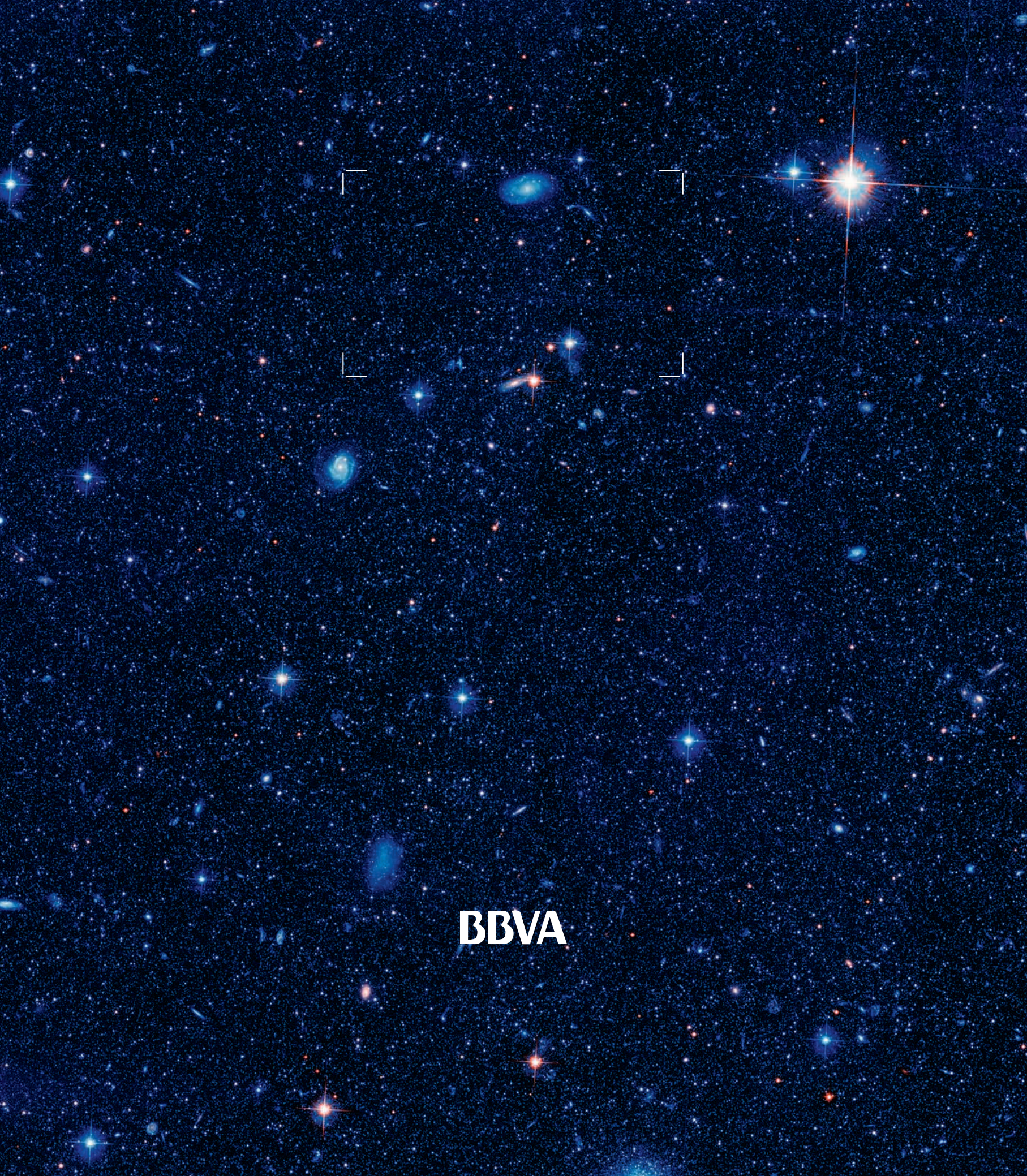
Encuadernación
Ramos

© BBVA, 2008
© de los textos, sus autores
© de las imágenes, sus autores

DL: M-54480-2008
ISBN: 978-84-935133-6-8

Fronteras del conocimiento se acabó de imprimir
en el mes de noviembre de 2008 en Madrid.
En su composición se utilizaron los tipos Rotis y Square.

La edición consta de 4.500 ejemplares
impresos en papel Gardapat de 150 gr.



BBVA