LA EVOLUCIÓN DE LA LIBERTAD

DANIEL C. DENNETT

Autor de *La conciencia explicada*

PAIDÓS TRANSICIONES

"El nuevo libro de Daniel Dennett combina, una vez más, un pensamiento filosófico original, una prosa maravillosamente viva y una argumentación extraordinariamente lúcida. *La evolución de la libertad* consigue lo que parecía imposible: decir algo nuevo sobre la libertad y el determinismo."

Richard Rorty

"Los intentos de encontrar un punto de vista útil sobre la libertad y el determinismo se han enfrentado siempre a toda clase de miedos y trampas conceptuales. [...] *La evolución de la libertad* es un libro que aclara enormemente las cuestiones evolutivas y cognitivas relacionadas con nuestra responsabilidad para tomar decisiones morales."

William H. Calvin

"Igual que otros fenómenos humanos, la libertad es el producto de la evolución. Daniel Dennett es un pionero en el estudio de esta evolución. La inteligencia artificial deberá tener en cuenta las ideas de Dennett en su intento de dotar de libertad a los robots."

John McCarthy

PAIDÓS TRANSICIONES

www.paidos.com

9 788449 315381 70051

DANIEL C. DENNETT

LA EVOLUCION DE LA LIBERTAD

PAIDOS
Barcelona
Buenos Aires
México

Título original: ***Freedom evolves***
Originalmente publicado en inglés, en 2003, por Viking Penguin, a member of Penguin Putnam Inc.

Traducción de Ramon Vila Vernis

Cubierta de Diego Feijóo

ISBN: 84-493-1538-7
Depósito legal: B-5.297/2004

Impreso en A & M Gráfic, S.L.
08130 Santa Perpetua de Mogoda (Barcelona)

Impreso en España - Printed in Spain

Para mi familia:
Susan, Peter, Andrea, Nathan y Brandon

Sumario

Prefacio

¿Cuánto tiempo he trabajado en este libro? Varias personas me hicieron esta pregunta durante la fase final de la edición y yo no sabía qué respuesta darles: ¿cinco años o treinta? Pienso que treinta años es una respuesta más cercana a la verdad, puesto que fue hace más o menos ese tiempo cuando comencé a reflexionar en serio sobre estos temas, a leer la literatura relevante, a esbozar argumentos, a preparar listas de libros y artículos para leer, a elaborar una estrategia y una estructura, y a entrar en debates y discusiones. Desde la perspectiva de conjunto que me dan estos treinta años, mi libro de 1984, *Elbow Room: The Varieties of Free Will Worth Wanting,* aparece como un proyecto piloto. El libro se basaba en un sencillo esbozo de diez páginas sobre la evolución de la conciencia (págs. 34-43), acompañado por dos notas a pie de página con dos promesas: al lector escéptico le debía sendas teorías debidamente detalladas sobre la conciencia y sobre la evolución. Me llevó doce años cumplir aquellas promesas, en *La conciencia explicada* (Dennett, 1991a) y *La peligrosa idea de Darwin* (Dennett, 1995). Durante todo ese tiempo no dejé de encontrar ejemplos del estado de cosas que inspiró y configuró *Elbow Room:* la agenda oculta que tiende a distorsionar la elaboración de teorías en el campo de las ciencias sociales y de la vida. Personas que trabajan en campos harto distintos, con diferentes métodos y programas de investigación, comparten muchas veces sin embargo una velada antipatía hacia las implicaciones de dos ideas, de las que tratan de mantenerse apartados: nuestras mentes no son otra cosa que lo que hacen nuestros cerebros, sin ninguna intervención milagrosa, y los talentos de nuestros cerebros son necesariamente el fruto de la evolución, igual que cualquier otra maravilla de la naturaleza. Sus esfuerzos por reprimir estas ideas bloquean su pensamiento y conceden un crédito indebido a dudosas versiones del absolutismo, y les incitan a ver grandes abismos donde hay pequeñas brechas fácilmente

sorteables. El objetivo de este libro es poner de manifiesto los espúreos edificios defensivos que se han construido para dar respuesta a este miedo, desmantelarlos, y reemplazarlos por unos mejores fundamentos para las cosas que más apreciamos.

En 2001, durante la fase de trabajo en casa, recibí ayudas muy valiosas, tanto institucionales como personales. La Universidad de Tufts, mi hogar académico de todos aquellos años, me concedió un semestre sabático. Una vez más la Villa Serbelloni de la Fundación Rockefeller, en Bellagio, me ofreció el entorno perfecto para escribir, y los primeros borradores de la mitad de los capítulos surgieron de un intenso mes de trabajo, iluminado por discusiones cargadas de sugerencias con otros residentes, sobre todo con Sheldon Siegel, Bernard Gross, Rita Charon, Frank Levy, Evelyn Fox Keller, Julie Barmazel, Mary Childers y Gerald Postema. Sandro Nannini y sus alumnos y colegas de la Universidad de Siena me proporcionaron una activa y entendida audiencia para la presentación de algunos de los argumentos centrales del libro.

En abril inicié un período de residencia como Profesor Visitante Leverhulme en la Escuela de Economía de Londres (LSE), donde presenté los primeros siete capítulos como conferencias públicas semanales, seguidas por seminarios celebrados al día siguiente, y complementadas por numerosos debates informales tanto en la LSE como en varias visitas a Oxford. John Worrall, Nick Humphrey, Richard Dawkins, John Maynard Smith, Matteo Mameli, Nicholas Maxwell, Oliver Curry, Helena Cronin, K. M. Dowding, Susan Blackmore, Antti Saaristo, Janne Mantykoski, Valerie Porter, Isabel Gois y Katrina Sifferd aportaron valiosas respuestas, refutaciones, matices y sugerencias.

Debo en buena medida a Christopher Taylor el cambio de perspectiva teórica propuesto en nuestro artículo conjunto y presentado en el capítulo 3, así como muchas sugerencias penetrantes a los borradores de otros s capítulos. Debo agradecer a David Benedictus, extraordinario escritor y amigo mío durante más de treinta años, varios cambios de perspectiva que llevaron finalmente a la elección del título. Robert Kane y Daniel Wegner, cuyos libros son objeto de crítica aquí (¡constructiva, espero!), fueron muy generosos en sus comentarios sobre mi tratamiento de sus obras. Entre los amigos y los compañeros que han leído amplias secciones de diversos borradores y me han dado consejo tanto sobre cuestiones de fondo como sobre asuntos relacionados con la edición cabe citar, en orden alfabético, a Andrew Brook, Michael Cappucci, Tom Clark, Mary Coleman,

Bo Dahlbom, Gary Drescher, Paulina Essunger, Marc Hauser, Erin Kelly, Kathrin Koslicki, Paul Oppenheim, Will Provine, Peter Reid, Don Ross, Scott Sehon, Mitch Silver, Elliott Sober, Matthew Stuart, Peter Suber, Jackie Taylor y Steve White.

Me mantuve fiel a mi tradición de jugar a Tom Sawyer y la valla pintada de blanco con el penúltimo borrador del libro, sobre el que se lanzaron durante mi seminario de otoño una nutrida e informada horda de estudiantes y revisores, tanto de cursos de posgrado como de licenciatura, hasta dejarlo inteligentemente reducido a piezas. James Arinello, David Baptista, Matt Bedoukian, Lindsay Beyerstein, Cinnamon Bidwell, Robert Briscoe, Hector Canseco, Russell Capone, Regina Chouza, Catherine Davis, Ashley de Marchena, Janelle DeWitt, Jason Disterhoft, Jennifer Durette, Gabrielle Jackson, Ann J. Johnson, Sarah Jurgensen, Tomasz Kozyra, Marcy Latta, Ryan Long, Gabriel Love, Carey Morewedge, Brett Mulder, Cathy Muller, Sebastian S. Reeve, Daniel Rosenberg, Amber Ross, George A. Samuel, Derek Sanger, Shorena Shaverdashvili, Mark Shwayder, Andrew Silver, Naomi Sleeper, Sara Smollett, Rodrigo Vanegas, Nick Wakeman, Jason Walker y Robert Woo aportaron comentarios que me permitieron introducir muchas mejoras. Los errores y las limitaciones que quedan no son culpa suya; pusieron lo mejor de su parte por corregirlos.

Debo también mi agradecimiento a Craig Garda y Durwood Marshall por las ilustraciones originales; a Teresa Salvato y Gabriel Love, del Centro de Estudios Cognitivos, por sus incontables rastreos bibliográficos y su ayuda con el papeleo durante la preparación de los muchos borradores del manuscrito; y al Collegium Budapest, que me ofreció una encantadora casa lejos de mi hogar que fue un gran estímulo intelectual para mí durante la fase final de correcciones y revisiones.

En último lugar, y el más importante de todos, quiero expresar una vez más mi amor y mi agradecimiento a mi esposa, Susan, por más de cuarenta años de consejo, amor y apoyo.

DANIEL DENNETT
20 de junio de 2002

Capítulo 1

Libertad natural

Una extendida tradición pretende que los seres humanos somos agentes responsables, capitanes de nuestro destino, *porque* en realidad somos *almas*, halos inmateriales e inmortales de material divino que habitan y controlan nuestros cuerpos materiales como unos titiriteros espectrales. Nuestras almas son la fuente de todo sentido, y el centro de todos nuestros sufrimientos, alegrías, glorias y vergüenzas. Pero la credibilidad de esta idea de las almas inmateriales, capaces de desafiar las leyes de la física, hace tiempo que quedó obsoleta gracias al avance de las ciencias naturales. Mucha gente piensa que este hecho tiene implicaciones terribles: en verdad no somos «libres» y nada importa realmente. El objetivo de este libro es mostrar en qué se equivocan.

DESCUBRIR LO QUE SOMOS

Si, abbiamo un anima. Ma é fatta di tanti piccoli robot.
Sí, tenemos un alma. Pero está hecha de muchos pequeños robots.

GIULIO GIORELLI

No es necesario que seamos almas inmateriales al estilo antiguo para estar a la altura de nuestras esperanzas; nuestras aspiraciones como seres morales cuyos actos y cuyas vidas importan no dependen *en absoluto* de si nuestras mentes obedecen o no a unas leyes físicas enteramente distintas del resto de la naturaleza. La imagen de nosotros mismos que podemos extraer de la ciencia puede ayudarnos a asentar nuestras vidas morales sobre nuevos y mejores fundamentos, y, una vez comprendamos en qué consiste nuestra libertad, estaremos en una posición mucho mejor

para protegerla frente a las amenazas genuinas que a menudo somos incapaces de reconocer.

Un alumno mío que ingresó en el Cuerpo de Paz para no prestar servicio en la guerra de Vietnam me contó más tarde sus esfuerzos para ayudar a una tribu que vivía en lo más hondo de la selva brasileña. Le pregunté si le habían exigido que les hablara del conflicto entre EE.UU. y la antigua URSS. Para nada, respondió. No hubiera tenido ningún sentido. No habían oído hablar nunca ni de Norteamérica ni de la Unión Soviética. En realidad, ¡no habían oído hablar nunca de Brasil! En la década de 1960 aún era posible que un ser humano viviera en un país, estuviera sujeto a sus leyes, y no tuviera la menor noción de ello. Si esto nos parece insólito es porque los seres humanos, a diferencia de todas las demás especies del planeta, somos, seres dotados de conocimiento. Somos los únicos que hemos comprendido lo que somos y el lugar que ocupamos en este gran universo. E incluso comenzamos a comprender cómo llegamos hasta aquí.

Estos descubrimientos más bien recientes sobre quiénes somos y cómo llegamos hasta aquí resultan, cuando menos, inquietantes. Somos un ensamblaje de unos cien billones de células de miles de tipos distintos. La mayor parte de estas células son «hijas» de la célula-óvulo y la célula-esperma, cuya unión dio inicio a nuestra existencia, pero en realidad se ven superadas en número por los billones de autoestopistas bacterianos de miles de cepas distintas almacenados en nuestro cuerpo (Hooper y otros, 1998). Cada una de nuestras células hospedadoras es un mecanismo inconsciente, un microrrobot en buena medida autónomo. No es más consciente de lo que puedan serlo sus invitados bacterianos. Ni una sola de las células que nos componen sabe quién somos, ni les importa.

Cada equipo de billones de robots está integrado en un sistema de eficiencia pasmosa que no está gobernado por ningún dictador, sino que se las arregla para organizarse solo para repeler a los extraños, desterrar a los débiles, aplicar sus férreas leyes de disciplina... y servir como cuartel general para un yo consciente, una mente. Tales comunidades de células son extremadamente fascistas, pero por fortuna nuestros intereses y nuestros valores tienen poco o nada que ver con los limitados objetivos de las células que nos componen. Algunas personas son amables y generosas, otras son despiadadas; algunas se dedican a la pornografía y otras consagran sus vidas al servicio de Dios. Durante largo tiempo ha sido muy tentador imaginar que tan notables diferencias debían obedecer a las propiedades especiales de un elemento *extra* (un alma) instalado de algún modo en el cuartel general del cuerpo. Ahora sabemos que por más tentadora que

siga resultando esta idea, no se ve apoyada por nada que hayamos aprendido acerca de nuestra biología en general y de nuestros cerebros en particular. Cuanto más aprendemos sobre cómo hemos evolucionado y sobre cómo funcionan nuestros cerebros, más seguros estamos de que no hay tal ingrediente extra. Cada uno de nosotros está *hecho* de robots inconscientes y nada más, sin ningún ingrediente no-físico, no-robótico. Todas las diferencias entre personas se deben a la forma en que se articulan sus particulares equipos robóticos a lo largo de toda una vida de crecimiento y experiencia. La diferencia entre hablar francés y hablar chino es una diferencia en la organización de las partes correspondientes, y lo mismo sucede con todas las demás diferencias de conocimiento y personalidad.

Puesto que yo soy consciente y usted es consciente, esos extraños y minúsculos componentes de los que estamos formados deben ser capaces de generar *de algún modo* un yo consciente. ¿Cómo es eso posible? Para comprender cómo es posible una obra de composición tan extraordinaria, debemos contemplar la historia de los procesos de diseño que hicieron posible la evolución de la conciencia humana. También debemos examinar de qué modo estas almas hechas de robots celulares pueden conferirnos los notables talentos y las obligaciones resultantes que supuestamente nos garantizaban las tradicionales almas inmateriales (por un proceso mágico no especificado). ¿Sale a cuenta cambiar un alma sobrenatural por un alma natural? ¿Qué perdemos y qué ganamos con el cambio? Algunas personas llegan a temibles conclusiones con relación a este punto, pero no hay ningún motivo para ello. Mi propósito es demostrar su error mediante una descripción de la aparición de la *libertad* en nuestro planeta, desde su origen hasta el nacimiento de la vida. ¿Qué clase de libertad? A medida que se desarrolle nuestra historia irán surgiendo diferentes tipos de libertad.

El planeta Tierra se formó hace cuatro mil quinientos millones de años, y al principio no contenía el menor rastro de vida. Y así permaneció durante tal vez quinientos millones de años, y luego, durante los siguientes tres mil millones de años más o menos, la vida se extendió por los océanos del planeta, pero era una vida ciega y sorda. Las células simples se multiplicaban, se engullían y se explotaban unas a otras de mil maneras distintas, pero no tenían ninguna noción del mundo más allá de sus membranas. Finalmente evolucionaron unas células mucho más grandes y complejas —las eucariotas—, todavía ciegas y robóticas, pero con la suficiente maquinaria interna como para comenzar a especializarse. Así siguieron las cosas durante algunos cientos de millones de años más, el tiempo que tardaron los algoritmos de la evolución en encontrar buenas formas para que esas

células y sus hijas y nietas se agruparan en organismos multicelulares compuestos por millones, miles de millones, y (finalmente) billones de células, cada una de las cuales cumple con una rutina mecánica concreta, pero ahora integrada en un servicio especializado, como parte de un ojo, una oreja, un pulmón o un riñon. Tales organismos (no los miembros individuales de los equipos que los componen) se habían convertido en seres capaces de conocer *a larga distancia*, capaces de espiar a su cena mientras trataban de pasar inadvertidos a media distancia, capaces de oír el peligro que les amenazaba desde lejos. Pero estos organismos completos todavía no sabían lo que eran. Sus instintos garantizaban que se aparearan con los organismos del tipo adecuado y formaran rebaño con los organismos del tipo adecuado, pero del mismo modo que aquellos brasileños no sabían que eran brasileños, ningún bisonte ha sabido nunca que era un bisonte.[1]

Sólo una especie, la nuestra, desarrolló evolutivamente otro truco: el lenguaje. Este ha supuesto para nosotros una autopista abierta hacia la posibilidad de compartir el conocimiento, en todos los órdenes. La conversación nos une, a pesar de nuestros distintos idiomas. Cualquiera de nosotros puede llegar a saber mucho sobre cómo es ser un pescador vietnamita o un taxista búlgaro, una monja de 80 años o un niño de 5 años ciego de nacimiento, un maestro del ajedrez o una prostituta. No importa lo distintos que seamos los unos de los otros, diseminados como estamos por todo el globo, pues podemos explorar nuestras diferencias y comunicarnos acerca de ellas. No importa lo parecidos que sean dos bisontes, plantados uno al lado del otro en medio de un rebaño, pues no pueden llegar a saber nada de sus parecidos, y no digamos ya de sus diferencias, porque no pueden comparar sus respectivas experiencias. Tal vez sean parecidas, pero son incapaces de compartirlas como lo hacemos nosotros.

Incluso en nuestra especie han hecho falta miles de años de comunicación para comenzar a descubrir las claves de nuestra identidad. Sólo hace unos cientos de años que sabemos que somos mamíferos, y hace sólo unas décadas que comprendemos con detalle cómo hemos evolucionado, junto con otros seres vivos, desde aquellos sencillos orígenes. Nos vemos supe-

1. En general, la naturaleza funciona de acuerdo con una versión del célebre Principio de Información Necesaria del mundo del espionaje: el bisonte no necesita saber que es un ungulado de la clase de los mamíferos —no podría hacer nada con dicha información, siendo un bisonte—; los brasileños no necesitaban saber demasiado (todavía) sobre el entorno más amplio del que formaba parte su entorno íntimamente conocido de la jungla, pero los brasileños, al ser seres humanos, podían extender sin demasiada dificultad sus horizontes epistémicos tan pronto como necesitaran saberlo. Estoy seguro que ahora ya lo saben.

rados en número en este planeta por nuestros primos lejanos, las hormigas, y apenas tenemos ninguna presencia al lado de otros parientes aún más lejanos, las bacterias. Aunque estemos en minoría, nuestra capacidad para compartir conocimiento a larga distancia nos da poderes que superan con mucho los de todos los demás seres vivos del planeta. Ahora, por primera vez en millones de años de historia, nuestro planeta está protegido por centinelas capaces de ver a gran distancia, capaces de anticipar peligros en un futuro lejano —un cometa en curso de colisión, o el calentamiento global— y diseñar planes para darles respuesta. El planeta ha desarrollado finalmente su sistema nervioso: nosotros.

Cabe la posibilidad de que no estemos a la altura de nuestra tarea. Cabe la posibilidad de que destruyamos el planeta en lugar de salvarlo, en gran medida porque somos librepensadores, creativos, exploradores y aventureros incontrolables, a diferencia de los billones de esclavos que nos componen. Los cerebros permiten anticipar el futuro, lo que da la posibilidad de corregir las acciones a tiempo para obtener mejores resultados, pero incluso la más inteligente de las bestias tiene horizontes temporales muy limitados y escasa capacidad, si es que tiene alguna, para imaginar mundos alternativos. En cambio nosotros, los seres humanos, hemos descubierto el don relativo de poder pensar incluso sobre nuestras propias muertes y más allá de ellas. Una gran porción de nuestro gasto de energía a lo largo de los últimos diez millones de años ha ido destinada a calmar las inquietudes que nos causa esta turbadora perspectiva de la que sólo nosotros disfrutamos.

Si uno quema más calorías de las que ingiere, no tarda en morir. Si descubre algunos trucos que le proporcionan un excedente de calorías, ¿en qué podría gastarlas? Tal vez podría consagrar siglos de esfuerzo humano a la construcción de templos, tumbas y piras sacrificiales sobre las que destruir algunas de sus más preciadas posesiones, e incluso algunos de sus propios hijos. Pero ¿por qué razón habría de querer hacer eso? Tan extraños y horribles dispendios nos dan algunas claves sobre los costes ocultos de los ampliados poderes de nuestra imaginación. No hemos accedido a nuestro conocimiento sin dolor.

¿Qué haremos a partir de ahora con nuestro conocimiento? Los dolores del parto de los nuevos descubrimientos aún no han remitido. Muchos temen que aprender demasiado sobre lo que somos —cambiar misterios por mecanismos— no hará sino empobrecer nuestra concepción de las posibilidades humanas. Es un miedo comprensible, pero si realmente corriéramos el riesgo de aprender demasiado, ¿acaso no serían los que están en la

punta de la lanza los que mostrarían mayores signos de inquietud? Echemos un vistazo a aquellos que participan en esta búsqueda de conocimiento científico y digieren con avidez los nuevos descubrimientos: es manifiesto que no les falta optimismo, convicción moral, entrega a la vida y compromiso con la sociedad. En realidad, si uno quiere encontrar ansiedad, desesperación y anomia entre los intelectuales de hoy, debe mirar hacia la tribu de moda en los últimos años, los posmodernos, a quienes les gusta proclamar que la ciencia moderna no es más que otro mito de una larga serie, y sus instituciones y su costoso equipo nada más que los rituales y los accesorios de una nueva religión. Que gente inteligente pueda tomarse en serio esta idea es un testimonio del poder que conserva el miedo ante las ideas, a pesar de nuestros avances en el autoconocimiento. Los posmodernos tienen razón en que la ciencia es sólo una de las cosas en las que podemos querer gastar nuestras calorías suplementarias. El hecho de que la ciencia haya sido la fuente principal de las mejoras en la eficiencia que han hecho posible este exceso de calorías no la hace merecedora de ninguna cuota especial de la riqueza que ha generado. Pero debería seguir siendo evidente que las innovaciones de la ciencia —no sólo sus microscopios, telescopios y ordenadores, sino su compromiso con la razón y la evidencia empírica— constituyen los nuevos órganos sensibles de nuestra especie, que nos permiten responder preguntas, resolver misterios y anticipar el futuro de modos a los que ninguna institución humana pretérita puede siquiera acercarse.

Cuanto más aprendemos sobre lo que somos, más opciones se abren ante nosotros a la hora de escoger lo que queremos ser. Los norteamericanos exaltan desde hace tiempo al hombre que se «hace a sí mismo», pero ahora que los nuevos conocimientos nos permiten rehacernos a nosotros mismos de maneras completamente nuevas, muchos se echan atrás. Muchos parecen preferir ir a tientas con los ojos cerrados, confiando en la tradición, antes que mirar a su alrededor para ver lo que va a ocurrir. Sí, es inquietante; sí, puede dar miedo. Al fin y al cabo, hay errores enteramente nuevos que por primera vez tenemos el poder de cometer. Pero es el comienzo de una gran aventura para nuestra especie. Y es mucho más excitante, y también mucho más seguro, ir con los ojos abiertos.

SOY QUIEN SOY

Recientemente leí en el periódico el caso de un padre joven que olvidó dejar a su hija pequeña en la guardería de camino hacia el trabajo. La niña

se pasó el día encerrada en el coche en un sofocante aparcamiento, y cuando por la tarde el padre pasó por la guardería le dijeron: «Hoy no vino a dejarla». El padre volvió corriendo al coche y la encontró todavía atada a su pequeño asiento en la parte de atrás, muerta. Traten de ponerse en la piel de aquel hombre, si son capaces de soportarlo. Cuando lo hago siento un escalofrío; mi corazón se encoge ante la idea de la vergüenza inconfesable, el autodesprecio, el arrepentimiento más allá de todo arrepentimiento posible que debe estar sufriendo ese hombre ahora. Y como persona notoriamente despistada, con tendencia a perderse fácilmente en sus pensamientos, me resulta especialmente inquietante hacerme la pregunta: ¿soy capaz de hacer algo así?, ¿podría ser tan negligente con la vida de un niño a mi cuidado? Reproduzco la escena con diversas variaciones, imagino las distracciones: un camión de bomberos que pasa a toda velocidad por mi lado justo cuando voy a girar hacia la guardería, algo que dicen en la radio y que me recuerda un problema que debo resolver aquel día y, más tarde, en el aparcamiento, un amigo que me pide ayuda al salir del coche, o tal vez unos papeles que se me caen al suelo y que tengo que recoger. ¿Es posible que se conjure una sucesión de factores de distracción de este tipo hasta hacerme olvidar el proyecto supremo de llevar a mi hija sana y salva hasta la guardería? Podría tener la mala suerte de encontrarme en una situación en la que los acontecimientos conspiraran para sacar lo peor de mí, poner en evidencia mis debilidades y arrastrarme por tan despreciable pendiente? Doy gracias de que no me haya encontrado aún en ninguna situación parecida, porque no estoy seguro de que no haya circunstancias en las que pudiera hacer lo mismo que hizo ese hombre. Cosas como ésas ocurren a cada momento. No sé nada más de ese padre. Es posible que sea un desalmado y un irresponsable, un villano que merece todo nuestro desprecio. Pero también es concebible que sea básicamente una buena persona, una víctima de una mala suerte cósmica. Y, por supuesto, cuanto mejor persona sea, mayor será el remordimiento que sentirá ahora. Debe de preguntarse si hay alguna forma honorable de continuar viviendo. «Soy el tipo que se olvidó a su hija pequeña en el coche y dejó que se cociera hasta la muerte en el vehículo cerrado. Ése soy yo.»

Cada uno es quien es, con sus verrugas y todo lo demás. No puedo convertirme en un campeón de golf o en un concertista de piano o en un físico cuántico. Puedo vivir con eso. Forma parte de quien soy. ¿Puedo bajar de los 90 en un circuito de golf o llegar a tocar esa fuga de Bach de principio a fin sin ningún error? Puedo probarlo, pero si nunca llego a tener éxito, ¿querrá decir eso que en realidad nunca podía haberlo hecho?

«¡Sé todo lo que puedas ser!» es un estimulante eslogan del ejército norteamericano, pero ¿no encierra una ridicula tautología? ¿Acaso no somos todos, automáticamente, todo lo que podemos ser? «Hola, soy un tipo gordo, ignorante e indisciplinado que por lo visto no tiene las agallas necesarias para ingresar en el ejército. ¡Ya *soy* todo lo que *yo* puedo ser! Soy quien soy.» ¿Se engaña a sí misma esta persona al negarse a probar una vida mejor, o ha llegado al fondo de la cuestión? ¿Hay algún sentido legítimo en el que aunque una persona *no* tenga ninguna posibilidad real y verdadera de ser un campeón de golf, *sí* tenga una posibilidad real y verdadera de bajar de los 90? ¿Puede alguno de nosotros hacer algo distinto de lo que termina haciendo? Y si no, ¿qué sentido tiene intentarlo? De hecho, ¿qué sentido tiene hacer nada?

Lo que queremos que se confirme a toda costa, de un modo u otro, es que nuestras acciones tienen sentido. Y durante varios milenios hemos luchado contra una familia de argumentos que apuntan en sentido contrario sobre la base de que si el mundo es tal como la ciencia nos dice que es, no hay espacio en él para nuestros empeños y aspiraciones. Tan pronto como los antiguos atomistas griegos soñaron la brillante idea de que el mundo estaba compuesto de una miríada de pequeñas partículas que chocaban unas contra otras, dieron con el corolario de que en tal caso todos los eventos, incluidos todos y cada uno de nuestros latidos, fibras y reflexiones privadas, se desarrollan de acuerdo con una serie de leyes de la naturaleza que *determinan* lo que va a ocurrir hasta sus más ínfimos detalles y no dejan, por lo tanto, ninguna opción abierta, ninguna alternativa real, ninguna oportunidad de que las cosas sean de un modo u otro. Si el *determinismo* es verdad, es una ilusión pensar que nuestras acciones tienen sentido, por más que *parezcan* tenerlo. En realidad, podemos poner todo el empeño que queramos en seguir pensando que lo tienen, pero con ello no haremos más que engañarnos a nosotros mismos. Esa es la conclusión que ha sacado mucha gente. Naturalmente, eso ha alimentado la esperanza de que después de todo las leyes de la naturaleza no sean deterministas. El primer intento de suavizar el golpe del atomismo vino de la mano de Epicuro y sus seguidores, quienes propusieron que una *desviación azarosa* en las trayectorias de algunos de esos átomos podía dejar espacio para la libertad de elección, pero como no tenían otra cosa que buenas intenciones para sustentar el postulado de esa desviación azarosa, la idea se encontró desde el principio con un merecido escepticismo. Pero no hay que abandonar la esperanza. ¡La física cuántica viene al rescate! Cuando descubrimos que en el extraño mundo de la física subatómica ri-

gen leyes distintas, leyes indeterministas, se abre un nuevo y legítimo campo de investigación: mostrar cómo podemos utilizar este indeterminismo cuántico para proponer un modelo del ser humano como agente que dispone de oportunidades genuinas y es capaz de tomar decisiones verdaderamente libres.

Se trata de una posibilidad de un atractivo tan perenne que merece un examen cuidadoso y considerado, y tal examen lo recibirá en el capítulo 4, pero mi conclusión final será, como han argumentado muchos antes que yo, que esa hipótesis no puede funcionar. Tal como lo expresó William James hace casi un siglo:

> **Si un acto «libre» es una novedad completa, la cual no proviene de mí, mi yo previo, sino *ex nihilo*, y simplemente se incorpora a mí, ¿cómo puedo yo, el yo previo, ser responsable de él? ¿Cómo puedo tener ningún carácter permanente que se mantenga lo suficiente como para merecer elogio o castigo? (James, 1907, pág. 53).**

¿Cómo es eso posible? Siempre aconsejo a mis estudiantes que estén atentos a las preguntas retóricas, pues marcan habitualmente la inferencia más débil en cualquier defensa. Una pregunta retórica supone un argumento por reducción al absurdo demasiado evidente como para que merezca ser formulado, el lugar perfecto para que se oculte una premisa incuestionada que podría revelarse merecedora de un rechazo explícito. A menudo se puede poner en una situación comprometida al que formula una pregunta retórica con el simple intento de responderla: «¡*Yo* te diré cómo!». En el capítulo 4 aplicaremos una estrategia de este tipo, y veremos que es posible responder en buena medida al reto de James. James va demasiado lejos en más de un sentido cuando concluye: «El rosario de mis días se deshace en un desorden de cuentas sueltas tan pronto como se retira la cuerda de la necesidad interna por influencia de la descabellada doctrina indeterminista». El indeterminismo no es descabellado, aunque tampoco será ninguna ayuda para aquellos que aspiran a la libertad, y nuestro examen pondrá al descubierto algunos pasos en falso que ha dado nuestra imaginación en su intento de encontrar una solución al problema de la libertad.

EL AIRE QUE RESPIRAMOS

La gente tiene una capacidad sorprendente para desviar su atención de las ideas que le resultan inquietantes, y en ningún caso se ha aplicado tanto

a ello como a la hora de pasar por alto el verdadero problema en la cuestión de la libertad. El problema clásico de la libertad, definido y desarrollado a lo largo de siglos de trabajo por parte de filósofos, teólogos y científicos, plantea la pregunta de si la constitución del mundo es tal que nos permite tomar decisiones genuinamente libres y responsables. La respuesta depende, según se ha pensado siempre, de algunos hechos básicos y eternos: las leyes fundamentales de la física (cualesquiera que sean) y ciertas verdades analíticas sobre la naturaleza de la materia, el tiempo y la causalidad, así como ciertas verdades igualmente analíticas y fundamentales sobre la naturaleza de nuestras mentes, como el hecho de que una piedra o un girasol no pueden ser libres en ningún caso, pues sólo algo dotado de mente puede ser candidato a este don, sea lo que sea. Trataré de demostrar que este problema tradicional de la libertad es, a pesar de su pedigrí, una cortina de humo, un acertijo de escasa importancia real que aparta nuestra atención de algunas preocupaciones vecinas que sí importan, que sí *deberían* tenernos despiertos toda la noche. Dichas preocupaciones suelen descartarse por considerarse complicaciones empíricas que enturbian las aguas metafísicas, pero mi intención es resistir a esta tendencia y promover dichas cuestiones tangenciales al rango de cuestiones principales. La amenaza genuina, la fuente oculta de toda la inquietud que convierte el tema de la libertad en un foco de atención tan perenne en los cursos de filosofía, surge de un conjunto de hechos relativos a la situación humana que son de naturaleza empírica, e incluso, en cierto sentido, política: son sensibles a las actitudes humanas. Tiene mucha importancia lo que pensemos sobre ellos.

Vivimos nuestras vidas sobre la base de ciertos hechos, algunos de ellos variables y otros sólidos como la roca. La estabilidad procede en parte de los hechos físicos fundamentales: la ley de la gravedad nunca nos abandonará (siempre tirará de nosotros hacia abajo mientras permanezcamos en la Tierra), y podemos confiar en que la velocidad de la luz se mantendrá constante hagamos lo que hagamos.[2] La estabilidad procede en parte también de otros hechos más fundamentales aún, de carácter *metafísico:* 2 + 2 siempre sumarán 4, el teorema de Pitágoras va a seguir siendo válido, y si A = B, todo lo que sea cierto de A es cierto de B y viceversa. La idea de que somos libres es otra condición de fondo para nuestro modo de pensar nuestras vidas. Contamos con ella; contamos con que la gente «es libre»

2. O casi constante. Ciertos datos recientes y controvertidos procedentes del espacio más remoto sugieren, en opinión de algunos científicos, que tal vez *podría* haber cierta variación en la velocidad de la luz a lo largo de períodos de tiempo cosmológicos.

del mismo modo que contamos con que caigan cuando los empujamos barranco abajo y con que necesiten comida y agua para vivir, pero en este caso no se trata ni de una condición metafísica de fondo ni de una condición física fundamental. La libertad es como el aire que respiramos, y está presente en casi todos nuestros proyectos, pero no sólo no es eterna, sino que es fruto de la evolución, y sigue evolucionando. La atmósfera de nuestro planeta evolucionó hace cientos de miles de años como resultado de las actividades de ciertas formas sencillas de vida terrestre, y continúa evolucionando hoy en respuesta a las actividades de los miles de millones de formas de vida más complejas cuya existencia ha hecho posible. La atmósfera de la libertad es otro tipo de entorno. Es una atmósfera que nos envuelve, nos abre posibilidades, configura nuestras vidas, una atmósfera *conceptual* de acciones intencionales, planes, esperanzas y promesas... y de culpas, resentimientos, castigos y honores. Todos crecemos en esta atmósfera conceptual, y aprendemos a conducir nuestras vidas en los términos que ella determina. *Parece* ser una construcción estable y ahistórica, tan eterna e inmutable como la aritmética, pero no lo es. Ha evolucionado como un producto reciente de las interacciones humanas, y algunas de las actividades humanas que se han desarrollado gracias a ella podrían amenazar también con perturbar su estabilidad futura, o incluso acelerar su desaparición. No hay garantía de que la atmósfera del planeta dure para siempre, como tampoco la hay de que lo haga nuestra libertad.

Ya estamos tomando medidas para evitar el deterioro del aire que respiramos. Tal vez no sean suficientes y tal vez lleguen demasiado tarde. Podemos imaginar ciertas innovaciones tecnológicas (grandes cúpulas acondicionadoras de aire, pulmones planetarios...) que nos permitirían vivir sin la atmósfera natural. La vida sería muy diferente, y muy difícil, pero seguiría valiendo la pena vivirla. ¿Qué ocurre, sin embargo, si tratamos de imaginar que vivimos en un mundo sin la atmósfera de la libertad? Sería vida, pero ¿seríamos *nosotros*? ¿Valdría la pena vivir la vida si dejáramos de creer en nuestra capacidad de tomar decisiones libres y responsables? Y ¿es posible que esta atmósfera omnipresente de la libertad en la que vivimos y actuamos no sea ni mucho menos un *hecho,* sino sólo una especie de fachada, una alucinación colectiva?

Hay quien dice que la libertad ha sido siempre una ilusión, un sueño precientífico del que apenas comenzamos a despertar. Nunca hemos sido *realmente* libres, y nunca podríamos haberlo sido. Pensar que hemos sido libres ha sido, en el mejor de los casos, una ideología que nos ha ayudado a configurar y a hacer más fáciles nuestras vidas, pero podemos

aprender a vivir sin ella. Algunas personas pretenden haberlo conseguido ya, pero no está claro a qué se refieren. Algunos insisten en que, aunque la libertad sea una ilusión, este descubrimiento no afecta para nada a su modo de pensar sobre sus vidas, sus esperanzas, proyectos y temores, aunque no se toman la molestia de desarrollar esta curiosa separación de conceptos. Otros excusan la persistencia de ciertos vestigios de aquel credo en sus formas de hablar y pensar diciendo que son hábitos básicamente inocuos que no se han tomado la molestia de superar, o que se trata de concesiones diplomáticas a las nociones tradicionales de los pensadores menos avanzados que les rodean. Siguen la corriente de la multitud, aceptan la «responsabilidad» por «decisiones» que en realidad no fueron libres, y culpan y elogian a los demás mientras cruzan los dedos bajo la mesa, pues saben que en el fondo nadie merece nunca nada, ya que todo lo que ocurre es simplemente el resultado de una vasta red de causas inconscientes que, en un último análisis, impiden que nada tenga ningún significado.

¿Están equivocados los que se autoproclaman desengañados? ¿Han abandonado una valiosa perspectiva sin ninguna buena razón para hacerlo, deslumhrados por una mala interpretación de la ciencia que les lleva a aceptar una imagen empobrecida de sí mismos? Y ¿tiene alguna importancia si es así? Resulta tentador desestimar toda la cuestión de la libertad como un acertijo filosófico más, un dilema artificial construido a partir de una conjunción de ingeniosas definiciones. ¿Eres libre? «Bien —dice el filósofo mientras enciende su pipa—, todo depende de lo que entiendas por libertad; ahora bien, por un lado, si adoptas una definición *compatibilista* de la libertad, entonces...» (y ya la tenemos montada). Para asegurarnos de que hay más en juego, de que estas cuestiones importan realmente, resulta útil trasladarlas al terreno personal. Así pues, reflexione usted sobre su vida adulta y escoja un momento realmente malo, un momento tan malo como sea capaz de contemplar con asfixiante detalle. (O, si eso resulta demasiado doloroso, trate simplemente de ponerse por un momento en la piel del joven padre.) Fije, pues, aquel acto terrible en su mente; fue usted quien lo hizo. ¡Qué no daría por no haberlo hecho!

¿Y qué? En el contexto general de las cosas, ¿qué sentido tiene su arrepentimiento? ¿Tiene algún valor, o es sólo una especie de hipo involuntario, un espasmo sin sentido causado por una palabra sin sentido? ¿Vivimos en un universo en el que tienen sentido la esperanza y el esfuerzo, el arrepentimiento, la culpa, la promesa, los propósitos de mejora, la condena y el elogio? ¿O forman parte todos de una vasta ilusión, reverenciada por la tradición pero cuyo tiempo ha pasado ya?

Algunas personas —tal vez sea usted una de ellas— encuentran un consuelo momentáneo en la conclusión de que no son libres, de que nada de todo eso importa, ni las faltas más denigrantes ni los triunfos más gloriosos; todo eso no es más que el despliegue de un mecanismo sin sentido. Pero aunque al principio les pueda parecer un gran alivio, tal vez luego se den cuenta, con irritación, de que a pesar de ello no pueden dejar de dar importancia a las cosas, no pueden evitar preocuparse, esforzarse, tener esperanzas... para luego darse cuenta de que no pueden dejar de sentir irritación ante este incesante deseo suyo de dar importancia a las cosas, y así sucesivamente, en una espiral descendente hacia el equivalente motivacional de la Muerte Caliente del Universo: nada se mueve, nada importa, nada.

Otras personas —tal vez sea usted una de ellas— están convencidas de que son libres. No sólo se plantean retos; se entregan a sus retos personales, desafían su supuesto destino. Imaginan posibilidades, tratan de sacar tanto como pueden de las oportunidades de oro que se les presentan y se estremecen cuando ven de cerca el desastre. Creen tener el control de sus propias vidas y ser responsables de sus propias acciones.

Habría, pues, según parece, dos tipos de personas: las que creen que no son libres (aunque no puedan evitar comportarse la mayor parte del tiempo como si creyeran serlo) y aquellas que creen ser libres (aunque sea una ilusión). ¿En qué grupo está usted? ¿A qué grupo le va mejor, cuál es más feliz? Y, en último término, ¿cuál tiene *razón*? ¿Acaso las del primer grupo son las que no se dejan engañar, las que ven más allá de la gran ilusión, al menos en sus momentos reflexivos? ¿O son ellas las que se engañan y son víctimas, por tanto, de ciertas ilusiones cognitivas que les llevan a la tentación de dar la espalda a la verdad, que les llevan a cerrarse posibilidades por descartar la idea misma que da sentido a sus vidas? (Es algo lamentable, pero tal vez no puedan evitarlo. ¡Tal vez estén *determinadas* a rechazar la idea de la libertad por su pasado, sus genes, su crianza, su educación! Como decía en broma el comediante Emo Phillips: «No soy un fatalista, pero aunque lo fuera, ¿qué podría hacer para evitarlo?».)

Esto plantea lo que podría ser otra posibilidad más. Tal vez haya dos tipos de personas normales (dejando a un lado las que están verdaderamente incapacitadas y no pueden en ningún caso ser libres porque están en coma o sufren un trastorno mental): aquellas que no creen en la libertad y *por ello mismo* no son libres, y aquellas que creen en la libertad y *por ello mismo son* libres. ¿Es posible que «el poder del pensamiento positivo», o algo por el estilo, sea lo bastante grande como para marcar la diferencia

crucial? Tal vez eso tampoco nos consuele demasiado, pues al parecer seguiría siendo cierto que sólo la suerte determina en qué grupo está cada uno, para bien o para mal. ¿Es posible cambiar de grupo? ¿Querría usted hacerlo? Es endiabladamente difícil mantener en perspectiva este curioso aspecto de la libertad. Si el hecho de que la gente sea (o no sea) libre es un hecho metafísico en bruto, entonces no puede verse influido por ninguna «ley de la mayoría» o nada por el estilo, y la única opción (¿opción?... ¿acaso seguimos teniendo *opciones*?) consiste en si queremos o no queremos saber la verdad metafísica, sea cual sea. Pero la gente habla y escribe a menudo como si, de hecho, estuvieran *haciendo campaña* en favor de la creencia en la libertad, como si la libertad (y no sólo la creencia en la libertad) fuera una condición política que pudiera estar bajo amenaza, que pudiera propagarse o extinguirse como resultado de lo que crea la gente. ¿Es posible que la libertad sea como la democracia? ¿Qué relación existe entre la libertad política y la libertad *metafísica* (a falta de mejor palabra)?

En lo que queda de libro, mi tarea será deshacer este nudo de ideas y ofrecer una visión unívoca, estable, coherente y empíricamente fundada de la libertad humana, y ya se sabe cuál es la conclusión a la que llegaré: la libertad es real, pero no es una condición previamente dada de nuestra existencia, como la ley de la gravedad. Tampoco es lo que la tradición pretende que es: un poder cuasi divino para eximirse del entramado de causas del mundo físico. Es una creación evolutiva de la actividad y las creencias humanas, y es tan real como las demás creaciones humanas, como la música o el dinero. Y es incluso más valiosa. Desde esta perspectiva evolutiva, el problema tradicional de la libertad se resuelve en una serie de cuestiones en buena medida por explorar, cada una de las cuales tiene su importancia a la hora de iluminar los problemas *serios* relacionados con la libertad; sin embargo, sólo podemos emprender este renovado examen una vez que hayamos corregido los errores en los que han caído los planteamientos tradicionales.

LA PLUMA MÁGICA DE ÜUMBO Y EL PELIGRO DE PAULINA

En *Dumbo,* la clásica película de dibujos animados de Walt Disney sobre un pequeño elefante que aprende a desplegar sus gigantescas orejas y volar, hay una escena clave en la que un indeciso —más bien aterrorizado— Dumbo escucha cómo sus amigos, los cuervos, tratan de convencerle para que salte al vacío y se demuestre a sí mismo que puede

volar. Uno de los cuervos tiene una idea brillante. Cuando Dumbo no mira le arranca una pluma de la cola a uno de los suyos y luego se la entrega ceremoniosamente a Dumbo, anunciando que es una pluma mágica: mientras Dumbo la lleve agarrada con la trompa, ¡podrá volar! La escena es presentada con magistral economía de medios. No se ofrece ninguna explicación, pues incluso los niños pequeños captan la idea sin necesidad de que se la digan: la pluma no es mágica en realidad, es una especie de amuleto que hará que Dumbo se eleve sobre el suelo por el poder del pensamiento positivo. Imaginemos ahora una variación de la escena. Imaginemos que otro de los cuervos, un escéptico de pueblo que es lo bastante listo como para captar el truco pero no como para comprender su virtud, se dispone a decirle a Dumbo la verdad cuando éste está inclinándose sobre el borde del precipicio, con la pluma firmemente agarrada. «¡Detengan a ese cuervo!», gritarían los niños. ¡Hagan callar a ese listillo, rápido, antes de que lo eche todo a perder!

A ojos de algunos, yo soy ese cuervo. Cuidado, avisan. Esta persona se dispone a hacer algo terrible, aunque sea con la mejor intención. Insiste en hablar sobre temas en los que es mejor no entrar. «¡Chist! Romperás el hechizo.» Esta advertencia no vale sólo para los cuentos de hadas; a veces resulta bastante apropiada para la vida real. Una docta disquisición sobre la biomecánica de la excitación sexual y la erección no es un buen tema de conversación para seducir a nadie, y las reflexiones sobre la utilidad social de la ropa y el ceremonial no son bienvenidos en un servicio fúnebre o un banquete de boda. Hay momentos en los que es más sabio apartar la atención de los detalles científicos, momentos en los que la ignorancia es fuente de felicidad. ¿Nos encontramos ante un caso de este tipo?

La capacidad de volar de Dumbo depende sólo circunstancialmente de la creencia de Dumbo en que puede volar. No es una verdad necesaria; si Dumbo fuera un pájaro (¡o sólo un elefante con más confianza en sí mismo!), su talento no sería tan frágil, pero siendo quien es, necesita todo el apoyo moral posible y no deberíamos permitir que nuestra curiosidad científica interfiriera en su delicado estado mental. ¿Es así también la libertad? ¿No parece al menos probable que nuestra libertad dependa de nuestra creencia en que somos libres? ¿Y no justifica este hecho que evitemos formular doctrinas que puedan minar esta creencia, con o sin razón? Aunque no podamos reírnos del chiste, ¿no estamos por lo menos obligados a cerrar la boca y cambiar de tema de conversación? Ciertamente hay quien piensa así.

En los muchos años que llevo trabajando en este problema, he terminado por reconocer una tendencia general. Mi punto de vista fundamental es el *naturalismo,* la idea de que las investigaciones filosóficas no son superiores ni previas a las investigaciones de las ciencias naturales, sino que van asociadas a dichos proyectos, y que el auténtico trabajo que deben hacer los filósofos en esta cuestión es clarificar y unificar las muchas perspectivas contrapuestas en una visión unificada del universo. Eso significa aceptar con gusto el tesoro de las teorías y los descubrimientos científicos que tanto esfuerzo ha costado reunir como material de base para las teorías filosóficas, de modo que sea posible una crítica recíproca informada y constructiva entre la ciencia y la filosofía. Cuando presento los resultados de mi naturalismo, mi teoría materialista de la conciencia (por ejemplo en *La conciencia explicada,* 1991a) y mi examen de los algoritmos darwinianos carentes de conciencia o finalidad que dieron origen a la biosfera y a todos sus productos derivados —tanto nuestros cerebros como nuestras ideas— (por ejemplo en *La peligrosa idea de Darwin,* 1995), me encuentro siempre con restos de incomodidad, un viento general de desaprobación o inquietud muy distinto del mero escepticismo. Habitualmente esta incomodidad se mantiene disimulada, como un leve rumor de un trueno lejano, una máxima de pensamiento positivo que distorsiona casi subliminalmente las actitudes. A menudo, después de que los interlocutores hayan agotado su repertorio de objeciones, alguien expone el motivo oculto de su escepticismo: «Todo eso está muy bien, pero ¿qué pasa entonces con la libertad? ¿No destruye su idea cualquier posibilidad de libertad?». Esta es siempre una respuesta bienvenida, pues apoya mi convicción de que la preocupación por la libertad es el motor que hay detrás de casi toda la resistencia al materialismo en general y al neodarwinismo en particular. Tom Wolfe, que está sin duda en perfecta sintonía con el espíritu de los tiempos, ha recogido esta idea en un texto que lleva el apremiante título de «Sorry but Your Soul Just Died» (Lo sentimos, pero su alma acaba de morir). Trata del surgimiento de lo que etiqueta de forma más o menos confusa como «neurociencia», cuyo ideólogo principal sería E. O. Wilson (el cual, por supuesto, no es ni mucho menos un neurocientífico, sino un entomólogo y un sociobiólogo), seguido de sus secuaces, Richard Dawkins y yo mismo. Para Wolfe la cosa está muy clara:

> **Como la conciencia y el pensamiento son productos del cerebro y el sistema nervioso enteramente físicos —y como el cerebro nace con todo graba-**

do—, ¿qué le hace pensar a usted que es libre? ¿De dónde podría venir su libertad? (Wolfe, 2000, pág. 97).

Yo tengo una respuesta. Simplemente, Wolfe se equivoca. Para empezar, el cerebro no «nace con todo grabado», aunque ése es sólo el menor de los malentendidos que hay detrás de esta extendida resistencia al naturalismo. El naturalismo no es ningún enemigo de la libertad; ofrece una explicación *positiva* de la libertad que da mejor respuesta a sus puntos oscuros que aquellas explicaciones que tratan de protegerla de las garras de la ciencia con una «oscura y miedosa metafísica» (en la acertada frase de P. F. Strawson). Presenté una versión de la misma en mi libro *Elboiv Room: The Varieties ofFree Will Worth Wanting,* de 1984. Pero a menudo me encuentro con que la gente duda de que pueda creer seriamente lo que digo. Están convencidos, como Tom Wolfe, de que el materialismo por definición no puede dejar espacio para la libertad y, mientras que Wolfe muestra al menos una actitud sarcàstica y distendida a propósito del tema («me encanta hablar con esa gente: manifiestan un determinismo incorruptible»), no es éste el caso de otros. Brian Appleyard, por ejemplo, ha escrito varias llamadas de alarma en forma de libros, aunque según otro alarmista, Leon Kass, él mismo podría haber caído también en la tentación:

A Appleyard no le gustan, y con razón, las implicaciones del pensamiento genocéntrico y manifiesta la esperanza de que pueda revelarse falso; en cualquier caso, insiste en que debemos resistimos a él. Pero no está filosóficamente preparado para mostrar dónde está el error. Peor aún, parece ser una víctima inconsciente de tal forma de pensar, haber tragado el anzuelo de los pomposos pronunciamientos de los más reduccionistas y grandiosos de los bioprofetas: Francis Crick, Richard Dawkins, Daniel Dennett, James Watson y E. O. Wilson (Kass, 1998, pág. 8).

Determinismo, genocentrísmo, reduccíonísmo... guardaos de esos pomposos bioprofetas; ¡pretenden subvertir todo lo que es más precioso en la vida! Frente a tan frecuentes condenas (y confusiones, como veremos), he sentido la necesidad de decir algo a modo de *apología.* ¿Me estoy comportando de modo irresponsable al preconizar tan activamente estas ideas?

Tradicionalmente, los sabios, en sus torres de marfil, no se han preocupado demasiado de la responsabilidad que les pueda corresponder por el *impacto ambiental* de su obra. Las leyes contra la difamación y la calumnia, por ejemplo, no eximen a nadie, pero fuera de estos casos la mayoría

de nosotros —incluidos los científicos de la mayoría de los campos— no acostumbramos a hacer declaraciones que puedan causar daño a otros, aunque sea indirectamente. Este hecho se manifiesta claramente en lo ridícula que nos parece la idea de un seguro profesional para críticos literarios, filósofos, matemáticos, historiadores y cosmólogos. ¿Qué podría hacer un matemático o un crítico literario, en el cumplimiento de sus deberes profesionales, para que pudiera necesitar la red de protección de un seguro profesional? Podría ponerle accidentalmente la zancadilla a un alumno en el corredor o se le podría caer un libro sobre la cabeza de alguien, pero aparte de estos daños colaterales más bien rebuscados, nuestras actividades son el paradigma de la inocuidad. O eso es lo que uno pensaría. Pero en aquellos campos en los que hay más en juego —y de forma más directa— existe una larga tradición que propugna la observación de una prudencia y un cuidado especiales para asegurar que no se produzca ningún daño (tal como profesa explícitamente el Juramento Hipocrático). Los ingenieros, conscientes de que la seguridad de miles de personas depende del puente que ellos diseñan, realizan pruebas especiales en condiciones predeterminadas dirigidas a garantizar la seguridad de sus diseños, de acuerdo con todos los conocimientos actuales. Cuando los académicos aspiramos a tener mayor impacto en el mundo «real» (y no sólo en el «académico»), debemos adoptar los mismos hábitos y actitudes que rigen en las disciplinas de orientación más práctica. Debemos asumir la responsabilidad de lo que decimos y reconocer que nuestras palabras, en caso de que alguien las crea, pueden tener profundos efectos, para bien o para mal.

No sólo eso. Debemos reconocer que nuestras palabras *pueden ser malinterpretadas* y que somos hasta cierto punto tan responsables de los malentendidos *probables* de lo que decimos como de los efectos «propios» de nuestras palabras. Se trata de un principio familiar: el ingeniero que diseña un producto potencialmente peligroso, en caso de uso indebido, es tan responsable de los efectos del uso indebido como de los efectos del uso debido, y debe hacer todo cuanto sea necesario para evitar usos peligrosos del producto por parte de personas inexpertas. Nuestra primera responsabilidad es decir la verdad hasta donde seamos capaces de hacerlo, pero no hay bastante con eso. La verdad puede ser dolorosa, sobre todo si la gente no la interpreta bien, y cualquier académico que piense que la verdad es una defensa suficiente para cualquier aserción seguramente no ha reflexionado lo suficiente sobre algunos de sus posibles efectos. En ocasiones, la probabilidad de que una *aserción verdadera* se malin-

terprete (o se haga de ella otro uso indebido) y el daño previsible que podrían causar tales malentendidos son tan grandes que hubiera sido mejor callarse.

Una antigua alumna mía, Paulina Essunger, propuso un vivido ejemplo para bajar las fantasías filosóficas a la fría realidad. Paulina había trabajado en el campo del sida y conocía bien los peligros que existen en este ámbito, por lo que llamaré a su ejemplo el Peligro de Paulina:

> **Pongamos por caso que yo «descubriera» que el sida puede ser erradicado de un individuo infectado bajo circunstancias ideales (total colaboración por parte del paciente, total ausencia de eventos que inhiban el efecto de la medicación, como náuseas, etc., total ausencia de contaminación con cepas diferentes del virus, etc.) tras cuatro años de seguimiento de un cierto régimen terapéutico. Puedo haberme equivocado. Puedo haberme equivocado en un sentido bastante simple y directo. Digamos que he calculado algo mal, que he leído mal ciertos datos, evaluando mal a los pacientes que han participado en el estudio o tal vez extrapolado demasiado generosamente.** *Aunque los resultados fueran ciertos, también podría hacer mal en publicarlos por el impacto ambiental que podrían tener.* **(Yendo más lejos, podría ser que los medios de comunicación hicieran mal en dar la noticia, o en darla de cierto** *modo.* **Pero parte de su responsabilidad parece revertir hacia mí. Especialmente si uso la palabra «erradicar», que en contextos virales se refiere habitualmente a eliminar el virus de la faz de la Tierra, no «meramente» liberar del virus a un paciente infectado.) Por ejemplo, podría extenderse una complacencia irracional entre los homosexuales masculinos: «Ahora el sida es curable, o sea que no tengo que preocuparme». Dicho relajamiento podría volver a disparar la incidencia de encuentros sexuales de alto riesgo dentro de este grupo. Es más, la extendida prescripción del tratamiento podría llevar a una rápida extensión de virus resistentes en la población infectada debido al periódico incumplimiento por parte de los pacientes (Essunger, correspondencia personal).**

En el peor de los casos, uno podría tener una cura para el sida, *saber* que tiene una cura para el sida y, sin embargo, no encontrar una manera responsable de hacer pública esta información. No sirve de nada indignarse con la complacencia o la temeridad de las comunidades de algo riesgo, como tampoco culpar a los pacientes faltos de voluntad que abandonan sus tratamientos a medio camino: todos esos son efectos predecibles y naturales (aunque lamentables) del impacto que tendría la publicación de la cura. Por supuesto, sería preciso explorar todas las vías prácticas para evitar los excesos derivados del descubrimiento y hacer planes para imple-

mentar todos los mecanismos de seguridad posibles, pero cabe la posibilidad de que, en el peor de los casos, fuera imposible alcanzar todos los beneficios imaginables del descubrimiento: simplemente no habría forma de llevarlos a la práctica. Esto no sería sólo un grave dilema; sería una tragedia. (De hecho, el caso hipotético de Paulina se está haciendo realidad ya en ciertos aspectos: el optimismo sobre una cura inminente ha llevado ya a actitudes peligrosamente relajadas en las prácticas sexuales de ciertos grupos de riesgo del mundo occidental.)

Así pues, esto entra en principio dentro de las posibilidades, pero ¿hay alguna probabilidad de que el intento de publicar una «cura» naturalista para el problema de la libertad pudiera enfrentarse a esta clase de obstáculos frustrantes? De hecho, existen unos cuantos obstáculos de este tipo, y no hay duda de que son frustrantes. Hay diversos guardianes del bien público que, con las mejores intenciones, quieren que *¡detengan a ese cuervo!* Están dispuestos a dar cuantos pasos sean necesarios para desanimar, silenciar o desacreditar a aquellos que en su opinión están rompiendo el hechizo antes de que el daño sea irreparable. Llevan años entregados a esa tarea y, aunque sus campañas han ido perdiendo crédito con la reiteración y sus colegas científicos han puesto en evidencia una y otra vez sus ostensibles falacias, los desechos de estas campañas no dejan de contaminar la atmósfera de los debates y distorsionan la comprensión que tiene el público general de estos temas. Los biólogos Richard Lewontin, León Kamin y Steven Rose, por ejemplo, dijeron en una ocasión que se veían a sí mismos como

> **una brigada de bomberos que no cesa de recibir avisos durante la noche para apagar el último fuego y responder a las emergencias inmediatas, pero que nunca tiene tiempo de elaborar planes para construir un edificio verdaderamente a prueba de incendios. Primero es la relación entre el CI y la raza, luego los genes criminales, luego la inferioridad biológica de las mujeres, luego la inmutabilidad genética de la naturaleza humana. Es preciso apagar todos estos fuegos deterministas con la fría agua de la razón para que no se incendie todo el vecindario intelectual (Lewontin y otros, 1984, pág. 265).**

Nadie ha dicho nunca que una brigada de bomberos tuviera que luchar con buenas artes, y esta brigada lanza mucho más que la fría agua de la razón sobre aquellos que ve como incendiarios. Y no son los únicos. Desde el otro extremo del espectro político, la derecha religiosa también domina el arte de la refutación por caricatura y se lanza sobre todas las oportunidades que encuentra para cambiar formulaciones cuidadosamen-

te matizadas de los hechos de la evolución por simplificaciones sensacionalistas frente a las que pueden luego escandalizarse y poner en guardia al mundo entero. Coincido con los críticos tanto de la izquierda como de la derecha en que se han producido *algunos* desafortunados pronunciamientos simplistas o exagerados por parte de algunas personas que han merecido sus críticas, y coincido en que tales faltas a la responsabilidad *pueden* tener efectos seriamente perniciosos. Es más, no cuestiono sus motivos, ni siquiera sus tácticas; si yo pensara que el mensaje que difunde cierta gente es tan peligroso que no puedo arriesgarme a concederle la audiencia justa, tendría cuando menos una fuerte tentación de distorsionarlo y caricaturizarlo en aras del bien público. Sentiría el impulso de inventarme algunos epítetos, como *determinista genético, reduccionista* o *fundamentalista darwinista*, y luego agitar esos espantajos ante el público. Como se acostumbra a decir, es un trabajo sucio, pero alguien tiene que hacerlo. Donde creo que se equivoca toda esta gente es al poner a naturalistas responsables y prudentes (como Crick y Watson, E. O. Wilson, Richard Dawkins, Steven Pinker y yo mismo) y a los escasos irresponsables que prefieren el sensacionalismo barato en el mismo saco y al atribuirnos ideas que nosotros hemos puesto gran cuidado en desautorizar y criticar. Como estrategia es inteligente: si realmente piensas que debes echar tierra encima de algo, usa una pala bien grande para estar seguro del resultado; ¡no dejes que los malos se oculten detrás de un escudo de rehenes respetables! Pero eso tiene el efecto de poner a algunos aliados naturales bajo fuego amigo, y es en último término una acción deshonesta, por decirlo claramente, da igual lo buenas que sean las intenciones.

El Peligro de Paulina al que nos enfrentamos los naturalistas es que siempre que formulamos versiones ponderadas y precisas de nuestras posiciones, algunos de esos guardianes del bien público aplican toda su inteligencia para transformar nuestras matizadas tesis en sonoras proclamas que resultan sin duda absurdas e irresponsables. He descubierto que cuanto más cuidado pongo en formular mi mensaje de forma clara y convincente, por ejemplo, más suspicaces se vuelven dichos guardianes. Parafraseando sus palabras, vienen a decir algo así como: «¡No prestes atención a todas las advertencias y complicaciones que despliega su embaucadora retórica! ¡Todo cuanto está diciendo, *en realidad*, es que no tienes conciencia, que no tienes mente, y que no eres libre! ¡No somos más que zombies y nada tiene ningún valor: eso es lo que dice *en realidad*\\». ¿Qué puedo responder a eso? (Por si acaso, que conste que no es eso lo que digo realmente.) Y para empeorar aún más las cosas, existen graves defec-

ciones y desacuerdos dentro de nuestro campo supuestamente monolítico de los «fundamentalistas darwinianos». Robert Wright, por ejemplo, cuyo reciente libro *Homero: The Logic of Human Destiny* es en la mayoría de los aspectos una excelente exposición de muchos de los temas que voy a presentar a continuación, se ve incapaz de suscribir la tesis central (según lo veo yo) de nuestra posición:

> **Por supuesto, el problema es aquí la tesis de que la conciencia es «idéntica» a los estados físicos cerebrales. Cuanto más se esfuerzan Dennett y otros por explicarme lo que quieren decir con eso, más me convenzo de que lo que realmente quieren decir es que la conciencia no existe (Wright, 2000, pág. 398).**

Tras varios cientos de excelentes páginas dedicadas a una concienzuda desmitificación naturalista, Wright vuelve, lamentablemente, a la visión mística de Teilhard de Chardin. (Una defección menos radical, pero más frustrante, es la de Steven Pinker [1997], cuyo constante coqueteo con las doctrinas mistéricas de la conciencia es en sí mismo otro misterio. Nadie es perfecto.)

Evidentemente, se trata de una cuestión que enciende las pasiones. Lo que tenemos aquí parece una carrera armamentista alrededor del evolucionismo, con una importante escalada por ambas partes. Pero nótese que en lugar de responder con el intento de ofrecer una caricatura aún mejor de mis oponentes, lo que preparo es un arma bien distinta en favor de nuestro bando: estoy tratando de despertar en ustedes la sospecha de que algunos de esos eminentes críticos puedan saber, en el fondo de su corazón, que tenemos razón. Al fin y al cabo el cuervo tenía razón, pero a pesar de ello siguen pensando: *¡detengan a ese cuervo!* Tal como veremos en capítulos posteriores, algunas de las objeciones más populares que se han planteado a la versión naturalista de la libertad se apoyan más en miedos que en razones. Los miedos son en sí mismos bastante razonables: si uno piensa que la caja que le ofrecen podría ser la caja de Pandora, es normal que multiplique sus recelos y agote todas sus objeciones antes de permitir que se abra la caja, pues entonces tal vez sea demasiado tarde.

¿Por qué persisto en mi intento de mantener mi punto de vista frente a tan encendida resistencia, sobre todo cuando reconozco que no puede descartarse del todo la posibilidad de que cause algún daño? (Los críticos exageran el peligro, por supuesto, al insistir en presentar estos puntos de vista en su versión más peligrosa; no hay duda de que juegan a hacerse las víctimas ante nosotros, los naturalistas.) Pues porque pienso que ya es

hora de que le quiten a Dumbo su pluma mágica. No la necesita, y cuanto antes lo aprenda mejor. En la película, recordarán ustedes, a Dumbo se le escapa la pluma en un momento crucial, en el que parece precipitarse hacia un final seguro, y no es hasta el último instante cuando comprende lo sucedido y extiende las orejas para salvarse. A esto se lo llama madurar, y pienso que estamos listos para hacerlo. ¿Por qué está mejor Dumbo sin su mito mágico? Pues porque su conocimiento de la realidad le hace ser menos dependiente, más autónomo y capaz. Trataré de demostrar que *algunas* de nuestras ideas tradicionales sobre la libertad están simplemente equivocadas; más aún, que son contraproducentes y ponen serios problemas al futuro de la libertad en este planeta. Por ejemplo, una comprensión realista de la libertad puede clarificar algunas de nuestras ideas sobre la culpa y el castigo, y calmar algunas de nuestras inquietudes respecto a lo que llamo el Espectro de la Exculpación. (¿Va a demostrar la ciencia que nadie merece un castigo? ¿O un elogio, por la misma razón?) También puede ayudar a reevaluar el papel que debe desempeñar la educación moral, y tal vez explicar incluso el importante papel que en el pasado han desempeñado las ideas religiosas para el sostenimiento de la moral dentro de la sociedad, un papel que ya no puede ser debidamente desempeñado por las ideas religiosas, pero que no podemos eliminar por completo sin correr un grave riesgo. Si nos aferramos a nuestros mitos, si no nos atrevemos a buscarles sustitutos científicamente contrastados, que ya tenemos a nuestra disposición, nuestros días de vuelo están contados. La verdad realmente os hará libres.

Capítulo 1

Una descripción naturalista de cómo hemos evolucionado nosotros y nuestras mentes parece amenazar el concepto tradicional de libertad, y el miedo ante esta perspectiva ha distorsionado la investigación científica y filosófica en esta materia. Algunos de los que han dado la voz de alarma ante los peligros de los nuevos descubrimientos sobre nosotros mismos han presentado una imagen muy falseada de los mismos. Una serena reflexión sobre las implicaciones de nuestro nuevo conocimiento sobre nuestros orígenes servirá de fundamento para una doctrina más sólida y prudente sobre la libertad que los mitos a los que está llamada a reemplazar.

Capítulo 2

Nuestra manera de pensar el deterninismo se ve a menudo distorsionada por ilusiones que pueden disiparse con la ayuda de una versión modelo, en la que puedan evolucionar entidades sencillas capaces de evitar el dolor y de reproducirse a sí mismas. Esto demuestra que el vínculo tradicional entre el determinismo y la inevitabilidad es un error, y que el concepto de inevitabilidad corresponde al nivel del diseño, no al nivel físico.

NOTAS SOBRE FUENTES Y LECTURAS COMPLEMENTARIAS

Las referencias completas de los libros y los artículos a los que se refiere el texto (por ejemplo, Wolfe, 2000) pueden encontrarse en la bibliografía que figura al final del libro. En cada capítulo ofreceré algunos comentarios suplementarios y remisiones a otras fuentes relacionadas con los temas objeto de discusión.

Es posible que a algunos lectores se les haya ocurrido pensar que el libro comienza con mal pie, pues caigo en una contradicción en la página 17. Primero niego que tengamos un alma distinta de los billones de células robóticas y luego observo como si tal cosa que somos conscientes: «Puesto que yo soy consciente y usted es consciente, esos extraños y minúsculos componentes de los que estamos formados deben ser capaces de generar *de algún modo* un yo consciente». Tal vez alguien se sienta tentado a decir, como Robert Wright, que mi tesis es en realidad que la conciencia no existe. Sería una lástima que tal convicción distorsionara su lectura del resto del libro, de modo que, por favor, traten de reservarse el juicio, ¡y dejen la puerta abierta a la posibilidad de que sea Wright quien se equivoque! Mi incorruptible materialismo es una parte integral de la perspectiva que pienso defender, y querría ser muy claro en este punto, aun a riesgo de generar antagonismo y escepticismo entre aquellos que añoran una concepción dualista de la conciencia. La articulación y defensa de esta teoría materialista de la conciencia puede encontrarse en mis libros antes mencionados, y aparece desarrollada y defendida contra varias críticas modernas en mis conferencias Jean Nicod, dictadas en noviembre de 2001 en París (en preparación d), así como en una serie de artículos publicados o de próxima publicación en varios libros y periódicos, y también disponibles en mi página web: http://ase.tufts.edu/cogstud

La literatura filosófica dedicada a la libertad es ingente, y en estas páginas sólo me centraré en una pequeña fracción de la obra reciente sobre el tema. Las obras que trataré contienen multitud de referencias que llevan a las demás. Al tiempo que yo introducía los últimos retoques a mi libro, se publicaron dos excelentes libros escritos por no-filósofos que cualquier persona interesada en el tema debería leer: *Breakdoivn of Witt,* de George Ainslie (2001), y *The lllusion of Conscious Witt,* de Daniel Wegner (2002). He introducido en mi libro unas breves reflexiones sobre estos dos textos, pero la riqueza de sus aportaciones va mucho más allá de lo que pueda deducirse de mis comentarios.

Capítulo 2

Una herramienta para pensar el determinismo

El determinismo consiste en la tesis de que «en cada momento dado hay exactamente un único futuro físicamente posible» (Van Inwagen, 1983, pág. 3). Podría pensarse que no es una idea particularmente difícil, pero sorprende ver cuán frecuentes son los errores sobre este punto, incluso entre autores muy rigurosos. En primer lugar, muchos pensadores asumen que el determinismo implica la inevitabilidad. No es así. En segundo lugar, muchos consideran evidente que el /«determinismo —la negación del determinismo— nos daría cierta libertad en cuanto agentes, cierta capacidad de maniobra, cierto margen, que simplemente no existiría en un universo determinista. Tampoco es así. En tercer lugar, se supone comúnmente que en un mundo determinista no hay *verdaderas* opciones, sino que éstas son sólo aparentes. Esto es falso. *¿En serio*? Acabo de contradecir tres ideas tan centrales en los debates sobre la libertad, y tan raramente cuestionadas, que muchos lectores deben suponer que estoy de broma, o que uso estas palabras en algún sentido esotérico. No es el caso; lo que digo es que la complacencia con la que acostumbran a tratarse estos temas, sin apenas entrar en ningún tipo de argumentación, es en sí misma un gran error.

ALGUNAS SIMPLIFICACIONES ÚTILES

Estas falsas nociones se encuentran en la base de todos los errores conceptuales relacionados con el libre albedrío y la libertad en general, de modo que, antes de poder hacer ningún avance en nuestra comprensión de cómo pudo evolucionar la libertad (en un universo que bien podría ser determinista), necesitamos equiparnos con algunos instrumentos de corrección, algunas herramientas para pensar que nos hagan menos vulnerables a los cantos de sirena de esas poderosas ilusiones. (Si siente usted aver-

sión hacia la argumentación filosófica sobre el determinismo, la causalidad, la posibilidad, la necesidad y el indeterminismo de la física cuántica, será mejor que pase directamente al capítulo 5, pero en tal caso debe usted renegar de cualquier dependencia intelectual respecto a aquellas tres proposiciones «evidentes», por más intuitivas que le parezcan, y aceptar como artículo de fe mis palabras cuando le aseguro que dichas proposiciones son errores que han llevado por el mal camino miles de debates sobre el tema. Sin embargo, casi puedo garantizarle que es imposible mantener dicha resolución, de modo que será mejor que se sumerja en mis demostraciones de estos tres errores, que tienen sus sorpresas y sus recompensas, y no presuponen ningún conocimiento previo de la materia.)

En la novela de Thomas Pynchon *El arco iris de la gravedad,* uno de los personajes pronuncia un notable discurso:

> **Pero habías caído en una ilusión más grande, y más peligrosa. La ilusión del control. Que A podía hacer B. Pero eso era falso. Completamente. Nadie puede *hacer* nada. Las cosas simplemente pasan (Pynchon, 1973, pág. 34).**

El personaje de Pynchon llega a la conclusión de que como los átomos no pueden *hacer* nada, y como las personas están hechas de átomos, en realidad las personas tampoco pueden *hacer* nada. No hay duda de que tiene razón cuando dice que existe una diferencia entre el hacer y el mero ocurrir, y también tiene razón cuando dice que nuestros intentos de comprender esta diferencia se ven amenazados por una peligrosa ilusión, pero interpreta esta ilusión al revés. El error no es tratar a las personas como si no estuvieran compuestas de multitud de átomos en relación con los cuales las cosas *ocurren* (como es el caso), sino *prácticamente* lo contrario: es tratar a los átomos como si fueran pequeñas personas que *hacen* cosas (lo que no es el caso). Ese error surge cuando extendemos las categorías que aplicamos a los agentes especialmente evolucionados al mundo más amplio de la física. El mundo en el que vivimos es el mundo de la *acción,* y cuando tratamos de imponer las nociones de este mundo al mundo de la física «inanimada» nos creamos un problema que fácilmente puede inducirnos a error.

La formulación adecuada de este aspecto de las complejas relaciones entre la física fundamental y la biología resulta intimidante, pero por fortuna existe una versión *modelo* de aquella relación que sirve perfectamente a nuestros propósitos. La diferencia entre un modelo y una herramienta desaparece si el modelo nos ayuda a comprender cosas que de otro modo nos resultarían demasiado complejas. La ciencia usa modelos a me-

nudo y saca grandes beneficios de ellos. Nadie ha visto nunca un átomo, pero todos sabemos el «aspecto» que tiene: un pequeño sistema solar, con un núcleo como un racimo apretado de uvas rodeado de electrones que orbitan en diferentes trayectorias con sus pequeños halos. Este viejo conocido, el modelo Bohr (figura 2.1), es sin duda una versión en gran medida simplificada y distorsionada, pero en muchos casos resulta de gran ayuda para comprender la estructura básica de la materia.

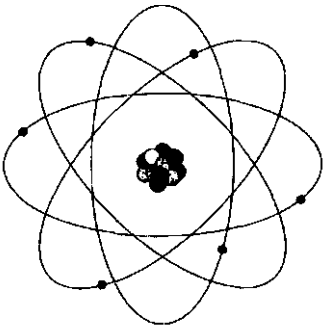

FIGURA 2.1. El átomo de Bohr.

Otro modelo cada vez más familiar para nuestra imaginación común es la gigantesca construcción de una doble hélice con multitud de escalones: el modelo Crick-Watson de la molécula de ADN (figura 2.2). También se trata de una útil simplificación.

FIGURA 2.2. La doble hélice del ADN.

Hace casi dos siglos el físico y matemático francés Pierre-Simon Laplace nos ofreció una imagen sencilla y vivida del determinismo, la cual ha orientado desde entonces nuestros modos de pensar y con ello también nuestras teorías y debates.

> Si hubiera un intelecto que en cualquier momento dado conociera todas las fuerzas que animan la Naturaleza y las posiciones respectivas de los seres que la integran, y fuera lo bastante vasto como para someter todos sus datos a análisis, podría condensar en una sencilla fórmula el movimiento tanto de los principales cuerpos del universo como el de sus átomos más pequeños: para un intelecto así no podría haber nada incierto; y el futuro estaría tan presente ante sus ojos como el pasado (Laplace, 1814).

Dadle a este intelecto omnisciente, a menudo llamado el ***demonio de Laplace***, una instantánea completa del «estado del universo» que muestre la localización exacta (y la trayectoria, masa y velocidad) de todas las partículas en aquel instante, y el demonio, con la ayuda de las leyes de la física, podrá anticipar cada colisión, cada rebote, cada roce que vaya a producirse en el instante siguiente, y actualizar la instantánea para ofrecer una nueva descripción del estado del universo, y así sucesivamente, hasta la eternidad.

En la figura 2.3, la instantánea se concentra en un momento *ti* y en sólo tres de los átomos del mundo, junto con sus diversas trayectorias, y el demonio usa esta información para predecir la colisión y el rebote de dos

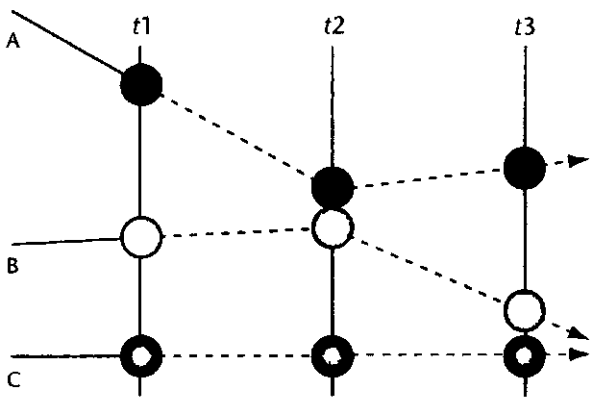

FIGURA 2.3. La instantánea de Laplace.

de ellos en *t2,* lo que lleva a nuevas posiciones en *t3,* y así sucesivamente. Un universo es *determinista* si existen reglas de transición (las leyes de la física) que *determinan exactamente* qué estado sucederá a cualquier estado particular que se describa. Si existe la menor desviación o incertidumbre, el universo es indeterminista.

En su formulación actual, este sencillo esquema contiene demasiados cabos sueltos: ¿cuan exacta debe ser la descripción de un estado?, ¿debemos realizar la proyección de todas las partículas atómicas?, ¿exactamente qué propiedades de las partículas debemos incluir en la descripción? Podemos resolver estas cuestiones mediante la adopción de otra idea simplificadora, la propuesta de W. V. O. Quine (1969) de restringir nuestra atención a universos imaginarios simples que él llama universos «democriteanos», en honor de Demócrito, el más inventivo de los antiguos atomistas griegos. Un universo democriteano consiste en unos cuantos «átomos» que se mueven por el «espacio». Eso es todo. Los átomos del universo democriteano no son átomos modernos llenos de complejidades cuánticas, sino verdaderos átomos *a-tómicos* (indivisibles, que no es posible partir), pequeños puntos uniformes de materia que no contienen partes, como los que postulaba Demócrito. El espacio que habitan debe ser también ultrasencillo, para lo cual debemos *digitalizarlo.* La pantalla del ordenador es un buen ejemplo de un *plano* digitalizado, una estructura bidimensional de cientos de líneas y columnas de pequeños *pixels,* o cuadros, cada uno de los cuales muestra en cada momento un color dentro de una serie finita de colores distintos. Si queremos digitalizar un espacio, un volumen tridimensional, necesitamos cubos *(vóxels,* en el lenguaje de los gráficos informáticos). Imaginemos un universo compuesto por una infinita cuadrícula de pequeños vóxels cúbicos, todos ellos completamente vacíos o completamente llenos (y conteniendo exactamente un átomo). Cada vóxel tiene una localización única o dirección en la cuadrícula, definida por sus tres coordenadas espaciales {*x, y,* 2}. Del mismo modo que cualquier sistema de gráficos informáticos en color dispone de una cierta gama de valores —diferentes tonos de color— que pueden aplicarse a cada píxel, en un universo democriteano todos los vóxels que no están vacíos (valor 0) contienen un átomo tomado de una gama limitada de átomos diferentes. Tal vez nos sea útil imaginarlos como diferentes colores: oro, plata, negro (carbón) y amarillo (sulfuro). Del mismo modo que podemos definir el conjunto de todas las imágenes posibles en una pantalla de ordenador (para cada sistema concreto de pixels y colores) como el conjunto de todas las permutaciones de las aplicaciones de los colores definidos a los pí-

xels, podemos definir el conjunto de todos los momentos del universo democriteano como el conjunto de todas las permutaciones de las aplicaciones de los diversos tipos de átomos a todos los vóxels.

Si lo que queremos es darle al demonio de Laplace una instantánea «completa» a partir de la cual trabajar, sabemos exactamente lo que debemos proporcionarle: la *descripción de un estado* dentro de un *universo democriteano,* que contenga el listado de los valores de todos los vóxels en un momento dado. Así, un fragmento de la descripción de estado D^ tendría el siguiente aspecto:

en el tiempo *t*\
vóxel {2, 6, 7) = plata,
vóxel {2, 6, 8) = oro,
vóxel {2, 6, 9} = 0,
... y así sucesivamente

No debemos preocuparnos por lo «detallada» que deba ser nuestra descripción, puesto que un universo democriteano tiene un límite definido, una diferencia mínima, y podemos comparar cualesquiera descripciones de estado del universo y descubrir las diferencias que pueda haber en el contenido de sus respectivos vóxels. Mientras haya un número finito de elementos distintos (oro, plata, carbón, sulfuro...) podemos ordenar todas las descripciones de estado —alfabéticamente, por ejemplo— en función del vóxel y el elemento que lo ocupa. La descripción de estado 1 es el universo vacío en el tiempo /; la descripción de estado 2 es igual que 1 excepto en que tiene un único átomo de aluminio que ocupa el vóxel (0, 0, 0); la descripción 3 traslada este solitario átomo de aluminio al vóxel {0, 0, 1); y así sucesivamente, hasta la última descripción (en orden alfabético), en la que el universo está lleno —todos los vóxels— de zinc. Incorporemos ahora el tiempo, la cuarta dimensión. Supongamos que en el siguiente «instante», el átomo de oro de {2, 6, 8} en D_k se mueve un vóxel hacia el este. Luego en Dk+i,

en el tiempo *t+1:*

vóxel {3, 6, 8} = oro.

Pensemos en cada «instante» de tiempo como un *frame* en una animación informática, que especifica el color o valor de cada vóxel en aquel

momento. Esta digitalización del espacio y el tiempo nos permite determinar las diferencias y los parecidos, y decir cuándo dos universos, o dos regiones o períodos de dichos universos, son exactamente iguales. Se puede resumir la historia de todo un universo democriteano con una sucesión de descripciones de estado, una para cada «instante», con independencia de lo que pueda durar dicho universo desde su Big Bang hasta su Muerte Caliente (o cualquier cosa que sustituya este principio y este final en tales mundos imaginarios). *En otras palabras, un universo democriteano es como un vídeo digitalde cierta duración, en tres dimensiones.* Podemos cortar el tiempo tan fino como queramos: treinta *frames* por segundo (como una película) o treinta billones de *frames* por segundo, dependiendo de cuáles sean nuestros propósitos. El tamaño de los vóxels es mínimo: un átomo indivisible por vóxel, no más. Quine propuso una simplificación ulterior: imaginemos que todos los átomos son iguales (como si fueran electrones), de modo que podamos tratar cada vóxel como vacío (valor = 0) o lleno (valor =1). Esta opción es como sustituir una pantalla en color por una pantalla en blanco y negro, una simplificación útil para ciertos propósitos, como veremos, pero no necesaria.

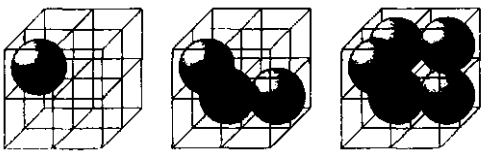

FIGURA 2.4. Tres de los 256 estados distintos de un universo democriteano de 8 vóxels.

¿Cuántas maneras distintas hay de llenar los vóxels con colores (o sólo con 0 y 1)? Por más que el tamaño de nuestro universo sea no sólo finito, sino positivamente pequeño, el número de posibilidades se multiplica muy rápidamente. Un universo que consista en sólo ocho vóxels (un cubo de dos por dos) y un solo tipo de átomo (vacío o lleno, 0 o 1), y que dure sólo 3 «instantes», tiene ya más de 16 millones de variaciones distintas ($2^{\cdot} = 256$ descripciones distintas, que pueden agruparse en 256^3 series distintas de tres). Un segundo de un universo contenido en un cubo de azúcar (al ritmo *lento* de 30 *frames* por segundo y suponiendo que el cubo tenga *sólo* un millón de átomos de ancho) daría un número de estados inimaginable.

En *La peligrosa idea de Darwin,* introduje el término *Vast* [«Vasto»] para referirme a aquellos números que, aunque finitos, son mucho más

grandes que las cantidades astronómicas. Lo usé para describir el número no infinito de libros que habría en la imaginaria Biblioteca de Babel de Borges, integrada por el conjunto de todos los libros posibles, y, por extensión, para describir el número de genomas que habría en la Biblioteca de Mendel, integrada por el conjunto de todos los genomas posibles. También acuñé el término *Vanishing* [«desVaneciente»] para definir, por ejemplo, el subconjunto de los libros *legibles*, casi inapreciable dentro de la Biblioteca de Babel. Llamemos al conjunto de todos los universos democriteanos posibles, es decir, al conjunto todas las combinaciones lógicamente posibles de átomos en el espacio y el tiempo, la Biblioteca de Demócrito. La Biblioteca de Demócrito es inconcebiblemente grande, por mucho que nos limitemos a un conjunto finito de parámetros (tipos de átomos, duraciones, etc.). Las cosas se ponen más interesantes cuando nos fijamos en subconjuntos concretos de la Biblioteca. Algunos universos de la Biblioteca de Demócrito están prácticamente vacíos, y otros están llenos de cosas; algunos experimentan gran cantidad de cambios a lo largo del tiempo y otros son estáticos (la misma descripción de estado, repetida eternamente). En algunos los cambios son completamente azarosos (en cada instante los átomos individuales aparecen y desaparecen como si fueran confeti); otros, en cambio, muestran pautas de regularidad y, por lo tanto, de predictibilidad. ¿Por qué algunos universos exhiben pautas? Simplemente porque la Biblioteca de Demócrito contiene todos los universos lógicamente posibles, de modo que en una u otra parte encontraremos *toda pauta posible-,* la única regla es que cada descripción de estado debe ser completa y consistente (sólo un átomo por vóxel).

En cuanto comenzamos a imponer reglas adicionales respecto a qué puede estar al lado de qué, y respecto a qué descripciones de estado distintas pueden ir después de otras en el tiempo, podemos llegar a subconjuntos más interesantes dentro de la Biblioteca. Por ejemplo, podemos prohibir la «aniquilación de la materia» por una regla que diga que cada átomo que existe en el tiempo *t* debe existir en algún sitio en el tiempo *t+1,* aunque puede moverse a otro vóxel si ese vóxel está vacío. Esto garantiza que el universo nunca pierda átomos con el paso del tiempo. (Dicho con mayor precisión, lo «prohibimos» al ignorar el Vasto número de universos que no obedecen a esta regla y limitar nuestra atención a los Vastos pero desvanecientes subconjuntos de aquellos que sí la obedecen: «Considera el conjunto C de universos en los que la siguiente regla se cumple siempre...».) Podríamos establecer un límite de velocidad (como por ejemplo la velocidad de la luz) al añadir que, entre un momento y el

siguiente, un átomo sólo puede moverse a un vóxel vecino, o podemos permitir saltos más largos. Podríamos decir que la materia *puede* ser aniquilada —o creada— bajo tales o cuales condiciones: por ejemplo, podemos tener la regla de que siempre que haya dos átomos de oro dispuestos uno encima del otro, en el instante siguiente desaparecerán, y en el vóxel inferior aparecerá un átomo de plata. Tales reglas de transición son el equivalente a las leyes fundamentales de la física para cada universo imaginario y es útil verlas como conjuntos de universos en los que se cumplen estas regularidades, con independencia de las otras diferencias que pudiera haber entre ellos. Supongamos, por ejemplo, que queremos «mantener una física constante» pero variar las «condiciones iniciales», es decir, el estado del universo en su momento inicial. Para ello debemos considerar el conjunto de universos en los que se cumpla siempre una determinada regla o conjunto de reglas de transición, pero donde las descripciones de estado de partida sean tan distintas como queramos. Esto viene a ser lo mismo que limitar nuestra atención, en la Biblioteca de Babel, a los libros escritos en inglés (gramatical); hay regularidades en la transición de carácter a carácter *(«i» antes de «e» excepto después de «c»...* y *Toda pregunta comienza con mayúscula y termina con signo de interrogación..),* pero los temas tratados cubren todas las variaciones posibles.

Una mejor analogía entre la Biblioteca de Babel de Borges y nuestra Biblioteca de Demócrito se basa en la existencia, dentro de la Biblioteca de Babel, de un Vasto número de libros que comienzan bien —por ejemplo novelas, libros de historia o de química— pero luego degeneran repentinamente en una ensalada sin sentido, un galimatías tipográfico. Por cada libro que puede leerse de cubierta a cubierta para gusto y provecho del lector, hay un Vasto número de volúmenes que comienzan bien, con las regularidades en la gramática, vocabulario, trama, desarrollo de los personajes y demás que constituyen los requisitos previos del *tener sentido,* y luego degeneran en una falta total de estructura. No hay ninguna garantía *lógica* de que un libro que comienza bien vaya a continuar bien. Lo mismo puede decirse de la Biblioteca de Demócrito. Tal era la idea de David Hume, ya en el siglo XVIII, cuando observó que por más que el sol se haya levantado cada día hasta ahora, *no hay contradicción* en suponer que mañana sucederá algo distinto, y que el sol no se levantará. Para expresar su observación en términos de la Biblioteca de Demócrito, digamos que existe un conjunto de universos, A, en los que el sol *siempre* se levanta, y un conjunto de universos, B, en los que el sol se levanta *hasta [digamos] el 17 de septiembre de 2004, en cuyo momento sucede algo distinto.* No hay

contradicción alguna en estos mundos: simplemente resulta que no «obedecen» a la misma física que se mantiene inmutable en los universos del conjunto A. La idea de Hume puede formularse del siguiente modo: por más hechos que podamos reunir sobre el pasado del universo donde nos encontramos, nunca podremos probar, desde un punto de vista lógico, que estamos en un universo del conjunto A, pues para cada universo del conjunto A existe un Vasto número de universos del conjunto B que son idénticos a él en cada vóxel/tiempo hasta el 17 de septiembre de 2004, y luego divergen de él en toda clase de sentidos sorprendentes o fatales.

Tal como señaló Hume, *esperamos* que la física que se ha cumplido hasta ahora en nuestro mundo siga cumpliéndose en el futuro, pero no podemos demostrar por pura lógica que seguiremos rigiéndonos por ella. Hemos llegado muy lejos en el descubrimiento de las regularidades que se han venido cumpliendo en el pasado en nuestro universo, e incluso hemos aprendido a hacer predicciones en tiempo real en relación con las estaciones, las mareas, la caída de los objetos, lo que uno encontrará si cava un agujero aquí o disecciona eso de allí, o bien si calienta esto y mezcla eso otro con agua, y otras cosas por el estilo. Tales transiciones son tan regulares, tan carentes de excepciones en nuestra experiencia, que hemos sido capaces de codificarlas y proyectarlas con la imaginación en el futuro. Hasta ahora nos ha salido bien; ha funcionado a las mil maravillas, pero no hay garantía lógica de que siga funcionando. Sin embargo, tenemos razones para creer que habitamos un universo donde este proceso de descubrimiento seguirá adelante *más o menos* indefinidamente, y producirá predicciones cada vez más específicas, fiables, detalladas y precisas, sobre la base de las regularidades que observamos. En otras palabras, podemos tomarnos a nosotros mismos como aproximaciones finitas e imperfectas del demonio de Laplace, pero no podemos demostrar, lógicamente, que nuestro éxito se va a mantener, sin presuponer las regularidades mismas cuya universalidad y eternidad pretendemos demostrar. Y existen otras razones, tal como veremos, para concluir que existen límites absolutos para nuestra capacidad de predecir el futuro. Si estos límites tienen alguna implicación respecto a nuestra autoimagen como agentes que toman decisiones y elecciones «libres», por las que se nos puede considerar propiamente responsables, es una de las resbaladizas cuestiones que vamos a tratar más adelante, y a la que nos vamos acercando con cautela, tratando de aclarar primero las cuestiones más sencillas. Nuestra forma de acercarnos gradualmente a nuestro tema, el *determinismo,* será ir definiendo un Vasto pero desVaneciente vecindario en el aún más Vasto espacio de los universos lógicamente posibles.

Algunos de los universos democriteanos tienen reglas de transición deterministas, y otros no. Consideremos el conjunto de los universos en los que especificamos que siempre que un átomo esté rodeado por vóxels vacíos tendrá una probabilidad de desaparecer de uno entre treinta y seis (en caso contrario, se mantiene en su sitio al instante siguiente). En tales universos es como si la Naturaleza lanzara un dado cada vez que uno de estos átomos se encuentra aislado de este modo; si sale un uno, el átomo «muere»; en cualquier otro caso, vive otro instante y la Naturaleza lanza otra vez el dado, a menos que el átomo tenga ahora un vecino. Esta sería una física *indeterminista*, que no especifica lo que ocurrirá a continuación en todos los aspectos, sino que deja algunas transiciones a la mera probabilidad. El demonio de Laplace debería esperar para ver cómo cae el dado antes de seguir adelante con su predicción del futuro. Otros conjuntos de universos obedecen a reglas de transición que no dejan nada al azar, que especifican exactamente qué vóxels están ocupados por qué átomos en el momento siguiente. Esos son los universos deterministas. Por supuesto, hay cuatrillones de variantes posibles en los universos democriteanos, algunas de ellas con reglas de transición deterministas y otras con reglas indeterministas.

¿Cómo podemos *saber* qué reglas de transición gobiernan un determinado universo democriteano? Podemos *estipular* una regla y luego considerar lo que podría o debería ser cierto si todos los miembros del conjunto obedecieran la regla, pero si nos dieran a estudiar un único universo democriteano, lo único que podríamos hacer es examinar la historia completa de sus vóxels y ver qué regularidades se mantienen, si es que se mantiene alguna. Podemos subdividir la tarea en sus partes naturales y buscar regularidades que se cumplan en los orígenes para luego ver si se siguen manteniendo durante todo el tiempo posterior. Aun teniendo presente el inquietante descubrimiento de Hume de que nunca podremos demostrar que el futuro será igual al pasado, podemos tratar de encontrar todas las regularidades que podamos y hacer la enorme pero tentadora apuesta —¿qué podemos perder?— de que el futuro *será* igual al pasado, de que no estamos en uno de esos extraños universos que al principio parecen ponérnoslo todo muy fácil, pero que se desmadran tras un período más o menos largo de regularidad.

Así pues, hemos encontrado una manera de dividir los universos democriteanos en deterministas, indeterministas y los que no son más que basura, que es como podríamos llamar a todos los universos *nihilistas* que no guardan ningún tipo de regularidad permanente en las transiciones. Nótese que, desde esta perspectiva, todo cuanto distingue el determinismo del indeterminismo es el hecho de exhibir un tipo u otro de regulari-

dad: sea una regularidad con probabilidades ineliminables menores a uno, o una regularidad en la que dichos factores probabilísticos están ausentes. No hay lugar, en otras palabras, para la pretensión de que dos universos democriteanos sean exactamente iguales en cada vóxel/tiempo, pero que uno de ellos sea determinista y el otro indeterminista.[1]

Queda clara pues la diferencia entre los universos democriteanos deterministas e indeterministas, pero para comprender exactamente lo que ésta significa (¡y lo que no significa!) lo mejor será mimar un poco más nuestras saturadas imaginaciones y considerar una imagen modelo aún más simplificada del determinismo. En primer lugar, cambiemos las tres dimensiones por dos (de vóxels a píxels), y concedámonos también la opción de Quine de que haya sólo blanco y negro, de modo que en cada momento dado un píxel esté simplemente en posición de ENCENDIDO o APAGADO. Acabamos de aterrizar en el plano donde se despliegan los impresionantes diseños del Juego de la Vida de Conway. Este audaz modelo extraordinariamente simplificado del determinismo fue desarrollado en la década de 1960 por el matemático británico John Horton Conway. La Vida de Conway ilustra vividamente las mismas ideas que nos interesan a nosotros sin necesidad de ningún conocimiento técnico de biología ni de física, ni de ningún conocimiento matemático más allá de la aritmética más sencilla.

DE LA FÍSICA AL DISEÑO EN EL MUNDO VLDA DE CONWAY

> *La complejidad de un ser vivo individual menos su capacidad de anticipación (en relación con su entorno) es igual a la incertidumbre del entorno menos su sensibilidad (con relación a ese ser vivo en particular).*
>
> JORGE WAGENSBERG, **«Complexity versus Uncertainty»**

1. En realidad, por definición, no puede haber *dos* universos democriteanos exactamente iguales en todos y cada uno de sus vóxel/tiempo. Una de las virtudes de la simplificación de Quine es que nos permite considerar los universos como si fueran *ediciones* de libros: cuando todos los elementos están en los mismos lugares y en los mismos momentos, existe una *identidad.* La propuesta de Quine para acotar nuestras reflexiones sobre los mundos posibles también pasa por alto la dudosa idea de que debamos conocer la *identidad* de los átomos individuales —no sólo su tipo: carbón u oro— para comparar el contenido de los vóxels entre un universo y otro. (Aviso para expertos: no entro a considerar la tradición de los mundos estándar posibles; también evito problemas familiares en relación con la identidad transmundana.)

Así pues, consideremos una cuadrícula bidimensional de píxels, cada uno de los cuales puede estar ENCENDIDO o APAGADO (lleno o vacío, blanco o negro).[2] Cada píxel tiene ocho vecinos: las cuatro celdas adyacentes (norte, sur, este y oeste) y las cuatro diagonales (noreste, sureste, suroeste y noroeste). El estado del mundo cambia con cada tictac del reloj en función de la siguiente regla:

> *Física de Vida*\ para cada celda de la cuadrícula, cuéntense cuántas de sus vecinas están ENCENDIDAS en el instante presente. Si la respuesta es que exactamente dos, la celda permanece en su estado actual (ENCENDIDA o APAGADA) en el instante siguiente. Si la respuesta es exactamente tres, la celda pasa a estar ENCENDIDA en el instante siguiente, con independencia de cuál sea su estado actual. En todas las demás condiciones la celda queda APAGADA.

Eso es todo. Esta única y sencilla regla de transición expresa toda la física del mundo Vida. Tal vez encuentre útil como truco mnemotécnico concebir esta curiosa física en términos biológicos: piense en las celdas que se encienden como si fueran nacimientos, las celdas que se apagan como si fueran muertes, y los instantes sucesivos como generaciones. La sobrepoblación (más de tres vecinos habitados) o el aislamiento (menos de dos vecinos habitados) lleva a la muerte. Pero recuerde, esto es sólo un truco para la imaginación: la regla dos-tres es la física básica del mundo Vida. Consideremos ahora cómo se desarrollan unas cuantas configuraciones sencillas de partida.

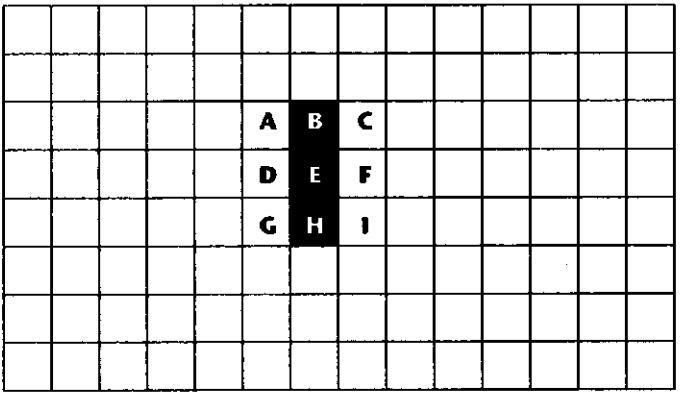

FIGURA 2.5. Luz intermitente en posición vertical.

2. Esta introducción a Vida está tomada, con algunas revisiones, de Dennett, 1991a, y Dennett, 1995.

Calculemos primero las celdas que nacen. En la configuración que aparece en la figura 2.5, únicamente las celdas *d* y /tienen exactamente tres vecinos ENCENDIDOS (celdas oscuras), luego serán las únicas celdas que nazcan en la siguiente generación. Las celdas *b* y *h* tienen sólo un vecino ENCENDIDO, de modo que mueren en la siguiente generación. La celda *e* tiene dos vecinos ENCENDIDOS, luego sigue adelante. Así pues, el siguiente instante tendrá el siguiente aspecto:

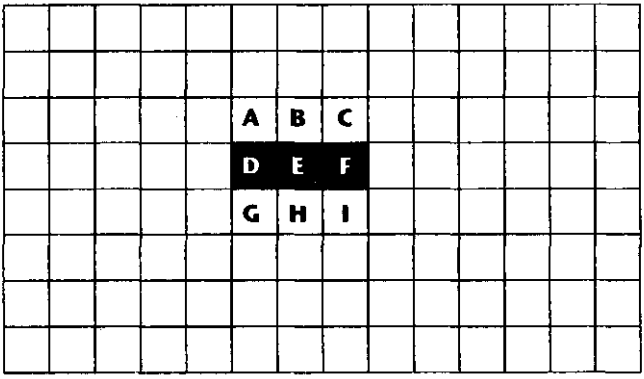

FIGURA 2.6. Luz intermitente en posición horizontal.

Obviamente, la configuración que aparece en la figura 2.6 revertirá de nuevo a su configuración anterior en el instante siguiente, y este pequeño diseño oscilará indefinidamente entre estas dos posiciones, a menos que de algún modo se introduzcan nuevas celdas ENCENDIDAS en la imagen. El diseño se llama luz intermitente o semáforo.

¿Qué sucederá con la configuración de la figura 2.7?

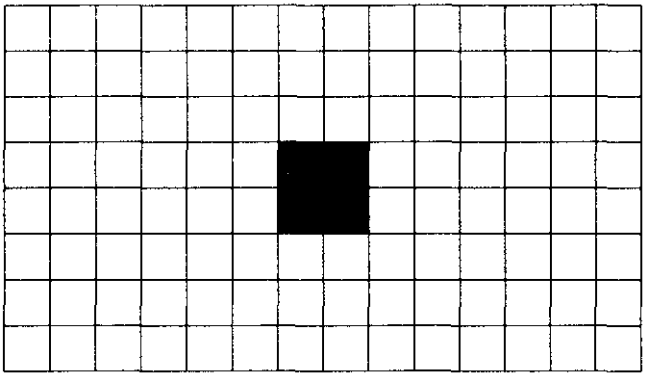

FIGURA 2.7. Naturaleza muerta cuadrada.

Nada. Cada celda ENCENDIDA tiene tres vecinos ENCENDIDOS, de modo que renacerá siempre tal como está. Ninguna celda APAGADA tiene tres vecinos ENCENDIDOS, de modo que no se producen nuevos nacimientos. Esta configuración se llama ***naturaleza muerta***; hay muchas otras configuraciones de tipo naturaleza muerta que no cambian con el tiempo.

Mediante una escrupulosa aplicación de nuestra sencilla ley, se puede predecir con perfecta precisión el instante posterior a cualquier configuración de celdas ENCENDIDAS y APAGADAS, y el instante posterior a éste, y así sucesivamente, ***de modo que cada mundo Vida es un universo democriteano bidimensional determinista.*** Y a primera vista, encaja perfectamente en nuestro estereotipo del determinismo: mecánico, repetitivo, ENCENDIDA, APAGADA, ENCENDIDA, APAGADA para toda la eternidad, sin que haya nunca una sorpresa, ni una oportunidad, ni una innovación. Si «rebobinamos» la cinta y comprobamos la secuela de cualquier configuración una y otra vez, siempre saldrá lo mismo. ¡Qué aburrido! ¡Gracias a Dios que no vivimos en un universo como éste!

Pero la primera impresión puede ser engañosa, sobre todo cuando estamos demasiado cerca de la novedad. Cuando damos un paso atrás y consideramos pautas más amplias de configuraciones de Vida, nos encontramos con algunas sorpresas. La luz intermitente tiene un período de dos generaciones que se repite ***ad infinitum,*** a menos que irrumpa alguna otra configuración. ***Son estas irrupciones las que hacen interesante la Vida.*** Entre las configuraciones periódicas hay algunas que nadan, como una ameba, por el plano. La más sencilla es el planeador, una configuración de cinco píxels (figura 2.8) a la que vemos en este caso dar un paso hacia el sureste.

Luego están los comilones, las locomotoras, los rastrillos y una multitud de otros habitantes certeramente bautizados del mundo Vida que emergen como objetos reconocibles a otro nivel. En cierto sentido, este nuevo nivel es simplemente una perspectiva a vista de pájaro del nivel bá-

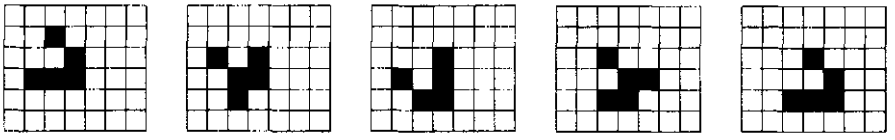

FIGURA 2.8. Planeador.

sico desde la que se perciben grandes formaciones de píxels en lugar de píxels individuales. Y es una satisfacción para mí decir que, cuando ascendemos a este nivel, llegamos a una instancia de lo que llamo el ***nivel del diseño***; un nivel que tiene su propio lenguaje, el cual ofrece una síntesis transparente de las tediosas descripciones que se podrían dar a ***nivel físico.*** Por ejemplo:

> Un comilón puede comerse un planeador en cuatro generaciones. Sea cual sea el objeto consumido, el proceso básico es el mismo. Se forma un puente entre el comilón y su presa. En la siguiente generación la región puente muere por sobrepoblación, llevándose consigo un mordisco tanto del comilón como de la presa. Luego el comilón se repara a sí mismo. La presa es normalmente incapaz de hacerlo. Si lo que queda de la presa muere, como sucede con el planeador, la presa ha sido consumida (Poundstone, 1985, pág. 38).

Nótese que ocurre algo curioso con nuestra «ontología» —nuestro catálogo de lo que existe— cuando cambiamos de nivel. En el nivel físico no hay movimiento, sólo **ENCENDIDO** y **APAGADO,** y los únicos individuos que existen, los píxels, se definen por una ubicación espacial fija, $\{x, y\}$. En el nivel del diseño nos encontramos repentinamente con objetos persistentes en movimiento; es uno y el mismo planeador (aunque compuesto en cada generación por píxels distintos) el que se ha movido hacia el sureste en la figura 2.8, cambiando de forma mientras se mueve; y hay un planeador menos en el mundo después de que haya sido eliminado por el comilón en la figura 2.9.

Nótese también que mientras en el nivel físico no hay excepciones de ningún tipo a la regla general, en el nivel del diseño nuestras generalizaciones deben ser acotadas: requieren cláusulas como «normalmente» («la presa normalmente es incapaz» de repararse a sí misma) o como «a menos que irrumpa alguna otra configuración». En este nivel, fragmentos sueltos de eventos anteriores pueden «romper» o «matar» a uno de los objetos de

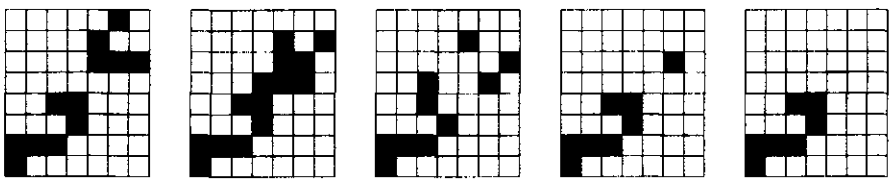

FIGURA 2.9. Comilón comiéndose a un planeador.

la ontología. Su *prominencia como objetos reales* es considerable, pero no está garantizada. Ha aparecido un elemento de mortalidad. Mientras que los átomos individuales —los píxels— entran y salen de la existencia, **ENCENDIDO** y **APAGADO**, sin posibilidad de acumular ningún cambio, ninguna historia que pudiera afectar a su historia posterior, las construcciones mayores sí pueden sufrir daños, experimentar una revisión de su estructura, o una pérdida o una ganancia de material que puede marcar la diferencia en el futuro. Las construcciones mayores también pueden mejorar, volverse *menos* vulnerables a una posible disolución posterior, por causa de algo que les haya ocurrido. Esta historicidad es la clave. La existencia en el mundo Vida de estructuras que pueden crecer, disminuir, torcerse, romperse, moverse... y en general persistir en el tiempo, abre las puertas a las oportunidades de diseño.

Hay una fraternidad de *hackers* de Vida a escala mundial que se ha lanzado a explorar estas oportunidades y que disfruta poniendo a prueba su ingenio en el diseño de configuraciones cada vez más elaboradas que hagan cosas interesantes en el plano Vida. (Si quiere usted explorar el mundo Vida, puede descargarse una excelente y accesible implementación Vida 32 de la página web http://psoup.math.wisc.edu/Life32.html. Tiene una biblioteca de configuraciones interesantes y enlaces con otras páginas. Siempre pido a mis alumnos que exploren el mundo Vida, porque he comprobado que da un sentido vivido y firme a un conjunto de intuiciones que de otro modo no existen, y les ayuda a pensar en estas cuestiones. En realidad —maravilla de las maravillas— a veces les hace *cambiar de idea* respecto a sus posiciones filosóficas. De modo que vaya con cuidado; puede ser una diversión adictiva... ¡y puede llevarle a abandonar su odio de toda la vida al determinismo!) Para convertirse en un *hacker* de Vida, sólo debe ascender al nivel del diseño, adoptar su ontología, y pasar a predecir —de manera imprecisa y asumiendo riesgos— el comportamiento de configuraciones mayores o sistemas de configuraciones, *sin molestarse en computar el nivel físico.* Puede imponerse la tarea de diseñar algún supersistema interesante a partir de las «partes» que el nivel del diseño pone a su disposición. Sólo lleva unos minutos pillarle el tranquillo, y quién sabe lo que puede ser usted capaz de crear. Por ejemplo, ¿qué pasaría si alineara a un grupo de comilones de naturalezas muertas, y luego los rociara con planeadores? Una vez que haya soñado su diseño, puede ponerlo a prueba inmediatamente; Vida 32 le informará rápidamente de cualquier problema que pudiera haber pasado por alto en sus predicciones sobre el comportamiento del diseño. Puede hacerse una idea de la riqueza de este

nivel del diseño a partir de unas citas que extraje de una excelente página web de Vida, http://www.cs.jhu.edu/~callahan/lifepage.html#newresults. La página web ya no está activa, lamentablemente, de modo que no hace falta que se moleste en encontrar el sentido a estos comentarios; simplemente pretenden ilustrar de qué modo piensan y hablan los *hackers* de Vida:

> **La hoja reacciona con toda la basura que produce R-pentomino mientras se transforma naturalmente en un Herschel, y milagrosamente reaparece algo más tarde sin dejar ningún residuo. Es necesario impedir que el primer planeador Herschel choque contra los últimos remanentes de la reacción, y no hay espacio para un comilón ordinario. Pero por suerte puede usarse en su lugar una bañera con cola y un bloque.**
>
> **Dave Buckinham encontró un reflector estable más rápido que no usa la reacción especial de Paul Callahan. En lugar de eso, el planeador que viene choca contra un barco para hacer un B-heptomino, que se convierte en un Herschel y es empujado hasta restaurar el barco. En este punto se necesita una forma compacta del Herschel de 119 pasos, tratándose de una naturaleza muerta no estándar, para hacer frente a la secuencia 64 64 77.**

Estos *hackers* de Vida juegan a ser Dios en sus universos simplificados en dos dimensiones, y se esfuerzan en diseñar configuraciones cada vez más impresionantes capaces de propagarse, transformarse, protegerse y moverse por el plano de Vida: en resumen, *hacer cosas* en el mundo, en lugar de limitarse a parpadear o, aún peor, persistir inmutables hasta la eternidad (a menos que irrumpa alguna otra configuración). Tal como revelan las citas, el problema al que se enfrenta cualquiera que juegue a ser Dios en este mundo es que, con independencia del buen aspecto que tenga la configuración inicial, siempre corre el riesgo de verse aniquilada, de convertirse en basura, o de ser comida por un comilón, o de desvanecerse sin dejar rastro.

Si usted quiere que sus creaciones persistan, deben estar protegidas. Si la física es constante (si no cambia la regla básica de Vida), la única cosa con la que puede jugar es con la descripción del estado inicial, pero ¡hay tantas entre las que escoger! Un conjunto de mundos de Vida de tan sólo un millón de píxels por un millón de píxels supone que el número de universos diferentes posibles por explorar asciende a 2 elevado a la billonésima potencia: la Biblioteca de Conway, una Vasta pero desVaneciente rama de la mucho más Vasta Biblioteca de Demócrito. Algunos de esos mundos de Vida son tremendamente interesantes, pero encontrarlos es

más difícil que encontrar una aguja en un pajar. La única forma de hacerlo, dado que es casi imposible que una búsqueda al azar dé resultados, es pensar en esta búsqueda como un problema de diseño: ¿cómo puedo *construir* una forma de Vida que *haga x* o *haga y* o *haga z?* Y una vez que he diseñado algo que puede *hacer x,* ¿cómo puedo proteger mi excelente *hacedor de x* de los daños que pueda sufrir después de construido? Después de todo, he dedicado una valiosa I+D (investigación y desarrollo) a diseñar mi *hacedor de x.* Sería una lástima que fuera destruido antes de que pudiera hacer su trabajo.

¿Cómo puede usted hacer cosas que duren en el mundo a veces hostil de Vida? Se trata de un problema objetivo, no antropomórfico. La física subyacente es la misma para todas las configuraciones de Vida, pero algunas de ellas, en virtud únicamente de su *forma,* tienen *capacidades* que otras configuraciones no tienen. Este es el hecho fundamental del nivel del diseño. Haga que sus configuraciones sean tan inhumanas, tan ajenas a nuestras categorías cognitivas, tan poco parecidas a un agente como sea capaz de imaginar. Si duran, ¿qué es lo que hay en ellas que explica este hecho? Una naturaleza muerta es magnífica hasta que algo se le echa encima. ¿Qué ocurre entonces? ¿Puede restaurarse a sí misma de algún modo? Tal vez sea mejor algo que pueda apartarse prestamente, pero ¿de dónde puede recibir el aviso de que vienen misiles? Tal vez sea aún mejor alguna cosa capaz de comerse los desperdicios que vengan y aprovecharse de ellos. La única regla es: cualquier cosa que funcione está bien. Bajo esta regla, lo que emerge a veces es sorprendentemente parecido a un agente, pero es probable que esto sea más el resultado de un sesgo de nuestra imaginación —como ver animales en las nubes sólo porque tenemos muchas «plantillas» de animales en nuestra memoria visual— que un hecho necesario. En cualquier caso, conocemos unos cuantos trucos que funcionan: un conjunto de trucos que recuerdan mucho nuestra propia biología. El físico Jorge Wagensberg ha sostenido que este parecido con la vida tal como la conocemos no es ningún accidente. En un artículo en el que no menciona la Vida de Conway, desarrolla una serie de definiciones de información, incertidumbre y complejidad a partir de las que obtiene algunas formas de medir la «independencia con respecto a la incertidumbre del entorno» y las utiliza para mostrar que la *persistencia,* o lo que él llama «mantener una identidad», en un entorno complejo depende (en términos de probabilidad) de varias formas de mantener esta «independencia», las cuales incluyen medidas «pasivas» como la simplificación (es el caso de las semillas y las esporas), la hibernación, el aislamiento (detrás de al-

guna barrera o en algún refugio) y el mero tamaño, y, por encima de todo, medidas «activas» que requieren capacidad de anticipación. «Una biota progresa en un entorno determinado si el nuevo estado de la biota es más independiente con relación a la incertidumbre del entorno» (Wagensberg 2000, pág. 504).

Un muro es a veces un buen negocio, si es lo bastante resistente como para que nada pueda romperlo. (¿Nada? Bueno, nada más pequeño que G, el proyectil más gigantesco que hemos lanzado contra él hasta el momento.) Un muro está simplemente ahí y encaja el golpe, sin *hacer* nada. Un protector móvil, por otro lado, debe moverse o bien en una trayectoria fija, como un centinela que hace la ronda alrededor del perímetro de un campamento, o bien en una trayectoria azarosa, limpiando los muros, o bien en una trayectoria guiada que dependa de la posibilidad de obtener alguna información sobre el entorno en el que se mueve. Un muro capaz de repararse a sí mismo es otra posibilidad interesante, pero es mucho más difícil de diseñar que una pared estática. Estos diseños más sofisticados, los diseños que pueden hacer cosas para mejorar sus opciones de supervivencia, pueden resultar bastante caros, ya que dependen de la posibilidad de reaccionar a informaciones acerca de las circunstancias. Su entorno *inmediato* (los ocho vecinos que rodean a cada píxel) resulta más que informativo: es completamente determinante; es «demasiado tarde para hacer nada» para evitar una colisión que ya ha comenzado. Si quiere que su creación sea capaz de *evitar* parte del daño inminente, deberá diseñarla para que o bien haga lo que debe hacer «automáticamente» (haciendo lo que hace siempre), o bien pueda anticiparlo de alguna manera, es decir, que pueda (ser diseñado para) guiarse por una u otra señal para adoptar una estrategia mejor.

Esto es el nacimiento de la evitación; el nacimiento de la prevención, la protección, la orientación, el desarrollo y todas las demás modalidades más sofisticadas y costosas de *acción.* Y justo en el momento de su nacimiento, podemos discernir una distinción clave que nos será de utilidad más adelante: algunos tipos de daños pueden, en principio, evitarse, y otros tipos de daños son *inevitables,* como se acostumbra a decir. La clave para evitar cosas es disponer de avisos previos, y éstos se ven estrictamente limitados en el mundo Vida por la «velocidad de la luz», que es (a todos los efectos prácticos) la velocidad a la que pueden desplazarse diagonalmente los patinadores simples a través del plano. En otras palabras, los planeadores podrían ser los *fotones,* las partículas de luz, en el conjunto de los universos Vida, y *reaccionar ante un planeador* podría ser una forma

de convertir una mera colisión o una irrupción en un modo de *informar,* el caso más sencillo de darse cuenta de algo o discriminar algo. Es fácil comprender por qué las calamidades que llegan a la velocidad de la luz «cogen por sorpresa» a cualesquiera creaciones con las que se encuentran; son realmente inevitables. Los problemas que se mueven a menor velocidad sí pueden, en principio, ser anticipados por cualquier forma de Vida que pueda extraer alguna información de la lluvia de planeadores (o de otras fuentes de información más lentas) que recibe y adoptar las medidas precisas. Tal vez pueda obtener información sobre lo que cabe esperar que ocurra de otras cosas que encuentre en su camino, pero sólo si *hay* alguna información en tales configuraciones que permita predecir la presencia de otras configuraciones en otros lugares o en otros momentos. En un entorno completamente caótico e impredecible, no hay ninguna esperanza de evitar nada como no sea por pura suerte.

Nótese que he estado mezclando dos procesos distintos para reunir información, y es importante que los distingamos con mayor claridad. En primer lugar, está la actividad de nuestros dioses-*hackers,* que están en condiciones de dirigir sus ojos y sus mentes hacia una gran variedad de mundos posibles Vida, y tratar de descubrir qué es lo que puede funcionar mejor, qué se revelará como resistente y qué como frágil. Por el momento, partimos del supuesto de que son realmente cuasi divinos en sus interacciones «milagrosas» con el mundo Vida: no están limitados por la lenta velocidad de los planeadores-luz; pueden intervenir, meterse dentro del plano y alterar el diseño de una creación en el momento que quieran, detener el mundo Vida a media colisión, deshacer el daño y volver al panel de dibujo para crear un nuevo diseño. Cada vez que *ellos* prevén una fuente de problemas pueden entregarse a la tarea de diseñar una manera de contrarrestarla. Sus creaciones serán los beneficiarios incautos e inconscientes de la previsión de los dioses-*hackers* que los han diseñado para que prosperen precisamente en aquellas circunstancias. Los dioses-*hackers* tienen sus limitaciones, sin embargo, y economizarán tanto como puedan. Por ejemplo, pueden estar interesados en cuestiones como: ¿cuál es la forma más pequeña de Vida capaz de protegerse de un daño *x* o un daño *y,* bajo las condiciones z (pero no bajo las condiciones *w)*? Después de todo, reunir información y ponerla en práctica es un proceso costoso y que consume tiempo, incluso para un dios-*hacker.* La segunda posibilidad es que los dioses-*hackers* diseñen configuraciones capaces de reunir *su propia* información a nivel local, dentro de los límites de la física del mundo que habitan. Es de esperar que cualquier creación finita que

use información será ahorrativa, y sólo conservará aquello que (probablemente) necesite o (probablemente) pueda usar, dadas las vicisitudes que se producen en su vecindario. Después de todo, el *dios-hacker* que la diseña quiere que sea lo bastante resistente como para cuidar de sí misma, no en todos los mundos Vida posibles, sino sólo en aquellos conjuntos de mundos Vida en los que tiene alguna probabilidad de encontrarse. Tal creación estará, en el mejor de los casos, en posición de *actuar como si supiera* que vive en un determinado *tipo* de vecindario, evitar un determinado *tipo* de daño o procurarse un determinado *tipo* de beneficio, pero no de actuar como si supiera exactamente en qué universo Vida habita.

Hablar de estos pequeños evitadores como si «supieran» algo supone una gran dosis de licencia poética, puesto que apenas podríamos imaginarnos algo más insensible a su entorno —son mucho más simples que una bacteria del mundo real, por ejemplo—, pero no deja de ser una buena manera de no perder de vista el trabajo de diseño que se ha invertido en ellos, gracias al cual poseen unas capacidades de *hacer cosas* de las que carecería cualquier agregación azarosa de píxels de más o menos el mismo tamaño. (Por supuesto, «en principio» —tal como a los filósofos les encanta decir— un Accidente Cósmico podría producir exactamente la misma constelación de píxels con exactamente las mismas capacidades, pero esto es una posibilidad enteramente negligible por su extrema improbabilidad. Sólo algo que haya costado un trabajo diseñar puede hacer cosas en un sentido interesante.)

Enriquecer la perspectiva del diseño hablando de las configuraciones como si «supieran» o «creyeran» algo y «quisieran» alcanzar un fin u otro supone pasar de la simple *perspectiva del diseño* a lo que llamo la *perspectiva intencional*. De acuerdo con ella pasamos a conceptualizar nuestros simples hacedores como *agentes racionales* o *sistemas intencionales*, lo cual nos permite pensarlos a un nivel aún más elevado de abstracción, e ignorar los detalles de cómo consiguen recoger la información en la que «creen» y cómo se las arreglan para «resolver» qué hacer, sobre la base de lo que «creen» y «quieren». Simplemente asumimos que sea cual sea su manera de hacerlo, lo hacen racionalmente, es decir, que sacan las conclusiones adecuadas sobre lo que deben hacer a partir de la información de la que disponen y en función de aquello que quieren. Eso le hace la vida mucho más fácil al diseñador de alto nivel, del mismo modo que conceptualizar a nuestros amigos y vecinos (y enemigos) como sistemas intencionales nos la hace mucho más fácil a nosotros.

Podemos ir saltando entre el punto de vista del dios-*hacker* y el «punto de vista» de las creaciones del dios-*hacker*. Los *áioses-hackers* tienen sus razones, buenas o malas, para diseñar sus creaciones tal como lo hacen. Tal vez las creaciones mismas ignoren cuáles son estas razones, pero *son* las razones por las que tienen las características que tienen, y si las creaciones persisten, es gracias a que tienen tales características. Si, más allá de eso, las creaciones han sido diseñadas para reunir información y usarla para guiar sus acciones, la situación se hace más complicada. La posibilidad más sencilla es que un dios-*hacker* haya diseñado un repertorio de reacciones-truco que tienden a funcionar en los entornos conocidos, algo análogo a los IRM (Innate Releasing Mechanisms [Mecanismos Desencadenantes Innatos]) y los FAP (Fixed Action Patterns [Patrones Fijos de Acción]) que los etólogos han identificado en muchos animales. Gary Drescher (1991) llama a esta arquitectura una *máquina de situación-acción* y la contrapone a la más compleja y costosa *máquina de elección*, en la que la creación individual genera sus propias razones para hacer *x o y* al anticipar los resultados probables de varias acciones posibles y evaluarlas en términos de los fines que también son fruto de su capacidad de representación (puesto que tales fines pueden cambiar con el tiempo, en respuesta a la nueva información reunida). Si preguntamos «en qué punto» las razones del diseñador dejan paso a las razones del agente diseñado, puede que descubramos que existe una gradación indiscernible de pasos intermedios, a lo largo de los cuales cada vez hay más trabajo del diseño que pasa del diseñador al agente diseñado. Una de las cosas más hermosas que tiene la perspectiva intencional es que nos permite ver claramente este cambio en la distribución de la «labor cognitiva» entre el proceso que origina el diseño y las actividades de la cosa diseñada.

Es posible que toda esta caprichosa manera de hablar sobre las configuraciones de píxels en Vida como si fueran agentes racionales le parezca a usted una exageración escandalosa, un burdo intento por mi parte de ponerle una venda en los ojos. Ha llegado el momento de hacer un test de sentido común: ¿exactamente cuánto puede *hacer,* en principio, una constelación de píxels de Vida, diseñada a partir de planeadores y otras configuraciones parecidas que vienen a ser como las «moléculas» del nivel del diseño, los bloques de construcción fundamentales para las formas de vida de nivel superior? Esta es la pregunta que inspiró inicialmente a Conway para crear el Juego de la Vida, y la respuesta que encontraron él y sus alumnos fue algo que nadie imaginaba. Pudieron demostrar que hay mundos Vida —esbozaron uno de ellos— en los cuales existe una Máqui-

na Universal de Turing, un ordenador bidimensional capaz en principio de computar cualquier función computable. Su tarea no fue precisamente fácil, pero demostraron que podían «construir» un ordenador operativo a partir de formas de Vida más sencillas. Corrientes de planeadores pueden proporcionar la «cinta» de entrada y salida de datos, por ejemplo, y el lector puede ser una gran asamblea de comilones, planeadores y otros fragmentos y piezas. El significado de todo esto es algo que cuesta de creer: cualquier programa que pueda ejecutarse en un ordenador podría, en principio, ejecutarse en el mundo Vida con una de estas Máquinas Universales de Turing. En el mundo Vida podría existir una versión del Lotus 1-2-3; y lo mismo podría decirse del Tetris o de cualquier otro videojuego. La capacidad de procesar información que tienen ciertas formas gigantescas de Vida es equivalente a la capacidad de procesar información de nuestros ordenadores reales en tres dimensiones. Cualquier competencia que se pueda «integrar en un chip» y montar en un artilugio en 3D puede ser imitada perfectamente por una constelación integrada de un modo parecido en una forma de Vida aún más grande en dos dimensiones. Sabemos que es algo que existe en principio. Todo cuanto hay que hacer es encontrarlo o, lo que es lo mismo, todo cuanto hay que hacer es diseñarlo.

¿PODEMOS REPRODUCIR EL *DEUS EX MACHINA?*

Ha llegado el momento de preguntarnos si deberíamos eliminar de nuestro cuadro a los dioses-*hackers* capaces de hacer milagros y reemplazar sus ingeniosos diseños por la evolución *dentro del propio mundo Vida.* ¿Hay algún mundo Vida, sea del tamaño que sea, donde el tipo de I+D que hemos venido describiendo hasta ahora pueda correr a cargo de la selección natural? Dicho con mayor precisión, ¿existen configuraciones del mundo Vida tales que, si se iniciara el mundo con una de ellas, ella misma haría *todo el trabajo* de los *¿ioses-hackers,* en el sentido de descubrir y propagar gradualmente configuraciones cada vez más aptas para la evitación? Este salto a una perspectiva evolutiva nos acerca a ideas familiares que parecen paradójicas o contradictorias desde nuestra perspectiva cotidiana, y se requiere un importante esfuerzo de reflexión para encontrarse cómodo con las transiciones entre ambas perspectivas. Uno de los primeros críticos de Darwin comprendió lo que se avecinaba y apenas pudo contener su indignación:

> **Según la teoría que se nos propone, la Ignorancia Absoluta es el artífice de todo; de modo que el principio fundamental de todo el sistema podría enunciarse como:** PARA HACER UNA HERMOSA Y PERFECTA MÁQUINA, NO ES NECESARIO SABER CÓMO HACERLA. **Un examen pormenorizado deja claro que esta proposición expresa, en forma condensada, el contenido esencial de la Teoría, y resume en unas pocas palabras todo lo que viene a decir el señor Darwin; el cual, por una extraña inversión del razonamiento, parece pensar que la Ignorancia Absoluta está perfectamente cualificada para ocupar el lugar de la Sabiduría Absoluta en el origen de todos los logros de la capacidad creadora (MacKenzie, 1868, pág. 217).**

MacKenzie identifica lo que llama una «extraña inversión del razonamiento», y no puede tener más razón. La revolución darwiniana es ciertamente una inversión del razonamiento cotidiano en más de un sentido, y resulta, por esta razón, extraña: un lenguaje *desconocido,* lleno de trampas para los incautos, por más que tengan una práctica considerable en su uso, tanto más porque muchos de sus términos son lo que los lingüistas llaman *falsos amigos,* términos que parecen ser sinónimos o compartir una misma raíz con términos de la propia lengua materna pero que difieren de ella de modos traicioneros. *One man's* Gift *is another man'spoison; one man's* chair *is another man's flesh* [«Lo que es un *regalo* para unos es un veneno para otros; lo que es una *silla* para unos es su propia carne para otros»] (Pista: consulte un diccionario alemán-inglés y otro francés-inglés.)* En el caso de la perspectiva darwiniana, el problema de los falsos amigos se ve exacerbado porque los términos que invitan a confusión se hallan estrechamente relacionados y son relevantes el uno para el otro (aunque no significan exactamente lo mismo). Cuando invertimos la tradicional perspectiva desde arriba y contemplamos la creación desde abajo, nos damos cuenta de que la inteligencia surge de la «inteligencia», de que la vista fue creada por un «relojero ciego», de que la elección emerge de la «elección», de que el voto deliberado surge del «voto» inconsciente, y así sucesivamente. Habrá muchas comillas en las explicaciones que están por venir. También descubriremos —¡hablando de paradojas!— que el todo puede ser más *libre* que las partes.

Vemos, pues, que la pregunta técnica inicial de si un proceso evolutivo podría sustituir la actividad de los dioses-*hackers* en el mundo Vida tiene algunas implicaciones de largo alcance. La respuesta a esta pregunta

*** Juego de palabras. El término inglés *Gift* («regalo») coincide con la palabra alemana para referirse al veneno, mientras que el término inglés *chair* («silla») coincide con la palabra francesa para referirse a la carne. (N. *del t.)***

tiene además algunos giros curiosos. En un mundo Vida de este tipo, debería haber entidades autorreproducibles, y ciertamente sabemos que existen, puesto que Conway y sus estudiantes integraron su Máquina Universal de Turing precisamente en una estructura de este tipo. El Juego de la Vida fue diseñado en realidad para explorar los experimentos mentales pioneros de John von Neumann acerca de los autómatas autorreproducibles, y sus creadores lograron diseñar una estructura autorreproducible capaz de poblar un plano vacío con tantas copias de sí mismo como fuera posible, de modo muy parecido a como se comporta una bacteria en una placa de petri, cada una de las cuales contendría una Máquina Universal de Turing. ¿Qué aspecto tendría esta máquina? Poundstone calcula que la construcción entera sería del orden de 10^{13} píxels.

> **Desplegar una configuración de 10" píxels requeriría una pantalla de vídeo de unos 3 millones de píxels de ancho como mínimo. Supongamos que los píxels sean de 1 milímetro cuadrado (lo que es una resolución' muy alta para los estándares de los ordenadores domésticos). En tal caso la pantalla debería tener 3 kilómetros (unas 2 millas) de ancho. Abarcaría un área unas seis veces superior a la de Monaco.**
>
> **La perspectiva empequeñecería hasta hacer invisibles los píxels de una configuración autorreproducible. Si nos pusiéramos lo bastante lejos de la pantalla para ver con comodidad toda la configuración, los píxels (e incluso los planeadores, los comilones y los cañones) serían demasiado pequeños para distinguirlos. Una configuración autorreproducible sería un halo brumoso, como una galaxia (Poundstone, 1985, págs. 227-228).**

En otras palabras, para cuando hubiéramos construido las suficientes piezas de algo capaz de reproducirse a sí mismo (en un mundo bidimensional), el resultado sería aproximadamente tanto más grande respecto a sus partes más pequeñas como lo es un organismo en relación con sus átomos. Eso no debería sorprendernos. Probablemente sea imposible hacerlo con algo menos complicado, aunque este extremo no ha sido demostrado en sentido estricto.

Sin embargo, la autorreproducción no es suficiente por sí misma. También necesitamos mutación, e incorporarla resultará sorprendentemente

3. Cuando Poundstone escribía (1985) ésa era una resolución muy alta, pero hoy resultaría baja. Los píxels de mi portátil son casi cuatro veces más pequeños, por lo que toda la pantalla a esta resolución ocuparía algo menos de 1 kilómetro de ancho, lo cual sigue siendo sin duda una pantalla grande.

costoso. En su libro *Le Ton Beau de Marot* (1997), Douglas Hofstadter llama la atención sobre el papel que desempeñan lo que llama las *intrusiones espontáneas* en cualquier proceso creativo, sea éste el resultado del esfuerzo de un artista humano, un inventor o un científico, o bien de la selección natural. Cada incremento del universo comienza por un momento de puro azar, la intersección imprevista de dos trayectorias que producen algo que resulta ser, visto en retrospectiva, como algo más que una mera colisión. Hemos visto cómo la detección de colisiones es una capacidad fundamental que puede ponerse al alcance de las formas de Vida y hasta qué punto las colisiones son uno de los principales problemas a los que se enfrentan todos los *hackers* de Vida, pero exactamente *¿cuántas colisiones* podemos permitirnos en nuestros mundos Vida? Eso se convierte en un grave problema cuando pretendemos añadir la mutación a las capacidades autorreplicadoras de las configuraciones de Vida.

Abundan las simulaciones informáticas de la evolución, y nos demuestran el poder de la selección natural para crear novedades sorprendentemente efectivas en períodos notablemente cortos de tiempo en este o aquel mundo virtual, pero su orden de magnitud es siempre, por fuerza, más limitado que el del mundo real, porque son mucho más *apacibles*. En un mundo virtual ocurre sólo lo que el diseñador especifica que ocurra. Consideremos una diferencia típica entre los mundos virtuales y los mundos reales: si nos ponemos a construir un hotel real, debemos dedicar gran cantidad de tiempo, energía y material para conseguir que aquellos que se encuentran en habitaciones contiguas no se oigan los unos a los otros; si nos ponemos a construir un hotel virtual, conseguimos el aislamiento gratis. En un hotel virtual, si queremos que las personas de las habitaciones contiguas puedan oírse unas a otras, debemos añadir esa posibilidad. Debemos añadir el *no*-aislamiento. Debemos añadir las sombras, los aromas, las vibraciones, la suciedad, las huellas y el desgaste. Todos estos rasgos no funcionales se dan gratuitamente en el mundo concreto y real, y tienen un papel crucial en la evolución. El carácter abierto de la evolución por selección natural depende de la extraordinaria riqueza del mundo real, que proporciona constantemente nuevos elementos *imprevistos* que pueden verse convertidos por azar, una vez cada tanto, en nuevos elementos de diseño. Para tomar el caso más sencillo, ¿puede haber suficientes interferencias en el mundo como para producir un número adecuado de mutaciones sin romper en el proceso todo el sistema reproductivo? En el sistema reproductivo de la Máquina Universal de Turing de Conway no había ruido, se producía cada vez una copia perfecta. La

mutación no estaba prevista en absoluto, con independencia del número de copias producidas. ¿Es posible diseñar un autómata autorreproducible aún más grande y ambicioso de tal modo que permita el impacto ocasional de algún planeador, como un rayo cósmico, capaz de producir una mutación en el código genético que está siendo copiado? ¿Es posible que un mundo Vida bidimensional pueda tener *ruido* suficiente como para hacer posible una evolución abierta y ser al mismo tiempo lo bastante *silencioso* como para permitir que las partes diseñadas puedan seguir haciendo su trabajo sin interferencias? Nadie lo sabe.

Resulta interesante el hecho de que para cuando hubiéramos especificado mundos Vida lo bastante complejos como para que fueran candidatos a tales potencialidades, serían demasiado complejos para que pudieran ser ejecutables en una simulación. Siempre puede añadirse ruido y basura al modelo, pero eso tiene el efecto de echar a perder la eficiencia misma que hace de los ordenadores unas herramientas tan magníficas. De modo que hay una especie de *homeostasis* o equilibrio autolimitador aquí. La misma simplicidad, el *exceso* de simplicidad de nuestros modelos evita que podamos modelar lo que más nos interesa, como es la creatividad, sea la de un artista humano o la de la propia selección natural, dado que en ambos casos esa creatividad se alimenta de la complejidad misma del mundo real. No hay nada misterioso o ni siquiera intrigante en esto, ningún rastro de extrañas fuerzas responsables de la complejidad o emergencias en principio impredecibles; no es más que un hecho práctico y cotidiano: el modelado informático de la creatividad se enfrenta a rendimientos decrecientes porque para conseguir una mayor apertura en el proceso debemos hacer el modelo más concreto. Debe modelar cada vez más colisiones incidentales como las que afectan a las cosas en el mundo real. En efecto, son las irrupciones las que hacen la vida interesante.

Así pues, es improbable que podamos demostrar jamás *por construcción* que en algún lugar de los Vastos confines del plano de la Vida haya configuraciones que imiten plenamente la apertura de la selección natural. Sin embargo, sí podemos construir algunas de las partes que integrarían una configuración de este tipo, lo que puede aportar importantes pruebas de su existencia. En efecto, existen configuraciones tales como las Máquinas Universales de Turing, y también objetos persistentes, capaces de protegerse y autorreplicarse, y procesos evolutivos limitados. Argumentos formales como el de Wagensberg (y también el de Conway, y el de Turing) nos permiten llenar los vacíos de lo que es imposible construir en la práctica, de modo que podemos decir con bastante seguridad que en nuestro mundo

determinista de juguete existen los ingredientes necesarios para la evolución de... *evitadores*\ Ésta es la proposición que necesitamos para disolver la ilusión cognitiva que asocia el determinismo con la inevitabilidad. Pero antes de llegar a esto, será útil volver del mundo de juguete al real, para ver qué es lo que sabemos sobre la evolución de la evitación en nuestro planeta.

DE LA EVITACIÓN A CÁMARA LENTA A LA GUERRA DE LAS GALAXIAS

Sabemos que en los primeros días —en los primeros miles de millones de años— de la vida en este planeta surgieron los diseños capaces de protegerse, gracias al lento y nada milagroso proceso de la selección natural. Hicieron falta del orden de mil millones de años de replicación de las formas de vida más sencillas para encontrar los mejores diseños —todavía susceptibles de revisión hoy, por supuesto— para los procesos básicos de replicación. En el camino hubo mucha *evitación* y *prevención,* pero a un ritmo demasiado lento para apreciarlas si no aceleramos el proceso artificialmente con la imaginación. Por ejemplo, el incansable proceso de exploración de la selección natural dio como resultado finalmente algunas secuencias de ADN contraproducentes, genes parásitos o *transposons,* que se colaban como polizones en los genomas de las anteriores formas de vida sin contribuir en nada al bienestar de dichas formas de vida: lo único que hacían era saturar sus genomas con nuevas copias (y copias de copias de copias) de sí mismas. Estos parásitos suponían un problema; debía de *hacerse* algo. Y a su debido momento el incesante proceso exploratorio de la selección natural, tras una búsqueda más o menos exhaustiva, «encontró» una solución (o dos, o más): diseños de estructuras en las partes valiosas y constructivas de los genomas que *evitaban* la proliferación excesiva de tales parásitos, *contrarrestando* sus *acciones* con *reacciones,* y así sucesivamente. Los genes parásitos reaccionaron a su vez ante este nuevo avance con un contraataque, desarrollado a lo largo de muchos cientos o miles de millones de generaciones, en un proceso incansable que continúa aún a día de hoy. En esta fase el límite de velocidad para la evitación no es la velocidad de la luz, sino la velocidad de la *generación.* El «acto» más sencillo de discriminación —simplemente «percibir» un nuevo problema y ponerse en posición para responder a él— requiere una generación, y el proceso de ensayo y error de «descubrir» una solución exige el sacrificio experimental de hordas de cepas mutantes a lo largo de muchas generaciones. Sin embargo, al final, los buenos diseños emergen victoriosos

(o bien la cepa desaparece, lo cual es el resultado más probable de todos esos «esfuerzos» de autopreservación). En el caso de algunas cepas afortunadas *ocurrió* que «encontraron» buenas respuestas. (No estaban *haciendo* nada, sólo eran parte de lo que estaba *ocurriendo,* la parte que tuvo la fortuna de nacer con mutaciones útiles en aquellas circunstancias.) Esos afortunados tuvieron descendientes, cuyos descendientes —los afortunados, otra vez— tuvieron descendientes, y así sucesivamente, hasta llegar a nosotros. Nosotros —los afortunados— estamos hechos de estas partes útiles, exquisitamente diseñadas para contribuir de forma útil a la evitación, pero ahora a una escala temporal mucho más rápida.

Y el proceso continúa hasta el presente. Matt Ridley describe el reciente y bien estudiado caso del llamado elemento P, un «gen saltador» que surgió en una cepa de laboratorio de moscas de la fruta (*Drosophila willistoni)* en la década de 1950, y luego se extendió a grandes poblaciones de sus primas, las *Drosophila melanogaster.*

> **El elemento P se ha propagado desde entonces como un incendio, hasta el punto de que la mayoría de las moscas de la fruta tienen el elemento P, aunque no aquellas atrapadas en el campo antes de 1950 y que se han mantenido aisladas desde entonces. El elemento P es un fragmento de ADN egoísta que demuestra su presencia por interferir en los genes entre los que salta. Gradualmente, los demás genes de la mosca de la fruta han aprendido a luchar contra él inventando formas de reprimir el hábito de saltar del elemento P (Matt Ridley, 1999, pág. 129).**

¿Cuánto tardaron estos genes en «reconocer» el problema y «luchar contra él»? Muchas generaciones, pero nótese que no hubo ninguna instancia central que lo reconociera, ni ninguna instancia decisoria. Lo que sucedió es exactamente lo que sucede siempre cuando opera la selección natural. El impacto de los elementos P no era uniforme en todas las cepas de la mosca de la fruta; existían variaciones en los genomas de estas moscas, algunas de las cuales estaban mejor preparadas para hacer frente a este nuevo reto. Aquellas que pudieron hacerle frente prosperaron, y aquellas de sus descendientes que demostraron estar aún mejor preparadas para hacerle frente prosperaron todavía más, de modo que a su debido tiempo surgieron «soluciones» al problema planteado por los elementos P, las cuales fueron «descubiertas» y «suscritas» por la Madre Naturaleza, también conocida como selección natural. Es un proceso que no puede ocurrir más rápidamente de como ocurre en la naturaleza; la experimentación no puede preceder al surgimiento del problema (eso sería precog-

nición evolutiva) y, por lo tanto, cada paso exige al menos una generación. Por fortuna, la experimentación puede aprovecharse de «desarrollos paralelos» si experimenta al mismo tiempo en todas las cepas actuales (aunque no en todas las posibles) de la mosca de la fruta, de modo que la resolución del problema puede avanzar con cierta prontitud, en menos de medio siglo en el caso de las moscas de la fruta.

Una de las advertencias habituales (y muy necesarias) que reciben los que estudian la evolución es la vieja sentencia sobre la falta total de previsión de la selección natural. No es que pretenda ponerla en duda. La evolución es el relojero ciego, y nunca debemos olvidarlo. Pero no deberíamos ignorar el hecho de que la Madre Naturaleza está bien provista de la sabiduría que da la perspectiva del tiempo. Su lema bien podría ser: «¿Si soy tan miope, cómo he llegado a ser tan rica?». Y si bien es cierto que la Madre Naturaleza carece de capacidad de previsión, también lo es que se las ha arreglado para crear seres —a nosotros los seres humanos, sobre todo— que sí disponen de ella y están comenzando a poner esta capacidad de previsión al servicio de guiar y corregir los procesos de selección natural en este planeta. De vez en cuando me encuentro con teóricos de la evolución, en ocasiones incluso muy brillantes, que encuentran paradójico este hecho. ¿Cómo es posible que un proceso sin capacidad de previsión invente un proceso que sí posea esta capacidad? Uno de los principales objetivos de mi libro *La peligrosa idea de Dartvin* era demostrar que esta idea no es en absoluto paradójica. El proceso de la selección natural, lento y carente de previsión, inventa procesos o fenómenos que aceleran el propio proceso de la evolución —grúas, no ganchos suspendidos del cielo, de acuerdo con mi terminología imaginaria— hasta que el trucado proceso evolutivo alcanza finalmente un punto en el que los experimentos que tienen lugar durante la vida de un individuo pueden afectar al lento proceso subyacente de evolución genética, e incluso, en ciertas circunstancias, tomar sus riendas.

Hoy los seres humanos somos capaces de ver y oír cosas a distancia, sin necesidad de esperar a tenerlas casi encima. Gracias a nuestros órganos de percepción a larga distancia y de las extensiones artificiales que hemos desarrollado para los mismos, podemos identificar y resolver problemas a un ritmo que se acerca al límite máximo de velocidad en el universo físico: la velocidad de la luz. Adelantarse a eso sería en todo caso precognición, de la que no disponemos, aunque por el momento ya le pisamos los talones a la velocidad de la luz en nuestras capacidades de reconocimiento y resolución de problemas. Gracias a nuestra tecnología, por ejemplo, podemos detectar el despegue de un misil nuclear sólo unos microsegundos después de que

éste tenga lugar a miles de kilómetros de distancia, y luego aprovechar este precioso tiempo para preparar alguna contramedida que tenga alguna posibilidad de éxito superior a cero. Se trata de un prodigio de la evitación, de la capacidad de esquivar un ladrillo que se acerca. (¿Podemos hacerlo realmente? ¿No he sostenido yo mismo que la Iniciativa de Defensa Estratégica de Ronald Reagan y sus derivados —conocidos en general como la guerra de las galaxias— son una fantasía tecnológica, de implementación completamente inviable? Pero si la guerra de las galaxias es imposible en la actualidad, tal como pienso en efecto que es, eso es sólo porque se sitúa en la vanguardia de la carrera armamentística de prevención *actual,* y las contramedidas imaginables parecen llevar ventaja; es casi seguro que otros conseguirían *prevenir* la *prevención,* que es el objetivo de la guerra de las galaxias, aunque es probable que muchos de los misiles lograran ser interceptados, que es todo cuanto afirmo aquí. No soy un fan de la guerra de las galaxias, a pesar de lo cual estoy encantado de ver que este sistema criminalmente irresponsable y caro pueda servir a fin de cuentas para algún modesto uso, ¡aunque sólo sea como el ejemplo de un filósofo!)

Hoy somos unos virtuosos de la evitación, la prevención, la interferencia y la anticipación. Hemos logrado llegar a la feliz situación de disponer del suficiente tiempo libre para examinar metódicamente el futuro y preguntarnos qué hacer a continuación. Exprimimos cada gota de información que podemos del mundo, y luego la moldeamos hasta construir asombrosas y novedosas perspectivas sobre lo que ha de venir. ¿Y qué es lo que vemos? Vemos algunas cosas *inevitables,* aunque nuestra lista se acorta cada semana que pasa. Antes no podíamos hacer nada para evitar los maremotos, o las epidemias de gripe, o los huracanes (todavía no podemos desviarlos, pero disponemos de las suficientes advertencias previas para que podamos ponernos a resguardo y minimizar los daños). Antes cuando una persona caía de un barco en plena noche en mitad del océano se la podía dar por perdida. Ahora podemos llevar helicópteros al lugar con sistemas de rastreo y sacar a la gente de las profundidades como ocurría en los milagros de pega del viejo *Deus ex Machina* de la tragedia griega. Todo esto es un desarrollo biológico muy reciente. Durante miles de millones de años no había nada parecido en este planeta. Los procesos eran enteramente ciegos o en el mejor de los casos miopes, insensibles y reactivos, nunca previsores y proactivos.

Tal como hemos visto, a nosotros nos resulta fácil reconocer, como los agentes inveterados e imaginativos que somos, la pauta de la prevención y la evitación a escalas temporales muy distintas, desde la supersónica hasta la superglacial. Podemos extenderla sin esfuerzo a los átomos e incluso a las

partículas subatómicas, y concebirlos, si nos gusta más así, como si fueran también pequeños agentes, preocupados por su futuro y ansiosos por contribuir a alguna gran campaña, mientras se esfuerzan por sobrevivir en un mundo hostil. Podemos imaginar, si así lo preferimos, que los átomos se encogen un momento antes de que se produzcan las colisiones previstas. Por supuesto, nada de eso tiene sentido. Los átomos no tienen capacidad de previsión, ni intereses, ni esperanzas; son sólo lugares minúsculos donde *ocurren* cosas, pero no se *hacen.* Pero eso no nos impide simplificar nuestra visión de ellos y tratarlos como si fueran agentes (aunque muy simples, con una idea única en la mente). Ese átomo de carbono se aferra tenazmente a esos dos átomos de oxígeno para *evitar* que se escapen y formar una molécula persistente de dióxido de carbono (una tarea modesta para un átomo de carbono). Otros átomos de carbono desempeñan papeles más excitantes al mantener juntas gigantescas proteínas formadas por gran cantidad de átomos, para que dichas proteínas puedan *cumplir con su trabajo*, sea cual sea.

Sospecho que nos parece natural contemplar las complejidades de los átomos y de los extraños habitantes del mundo de la física subatómica y tratarlos como si fueran pequeños agentes porque nuestros cerebros están diseñados para tratar todo lo que encontramos como un agente siempre que sea posible (por si acaso lo fuera realmente). En los primeros días de la cultura humana, en la infancia de la civilización, podríamos decir, encontrábamos útil abusar de este *animismo* y tratar el conjunto de la naturaleza y tratarlos como si estuviera hecho de dioses y hadas, espíritus benevolentes y malvados, duendes y trasgos, los cuales serían los responsables de todos los fenómenos naturales que observábamos. Podría decirse que todo eran sistemas intencionales. Esta táctica se ha moderado y sofisticado desde entonces —desde Demócrito, de hecho—, de modo que ahora nos encontramos bastante cómodos pensando en los átomos como pequeñas partículas inconscientes que rebotan de un lugar a otro. No *actúan* exactamente, pero a pesar de todo *hacen cosas-,* repeler y atraer, detenerse por un momento en un sitio o salir disparados de él.

No estoy sugiriendo que se pueda establecer una distinción transparente, en último término, entre las cosas que meramente ocurren y las cosas que hacen cosas, por más valiosa que sea esta oposición. Como de costumbre, lo que tenemos es una gradación que va de los colores más chillones a los tonos pastel y finalmente a lo invisible, una gradación en la que cada vez resulta menos apropiada la familia de conceptos asociados a nuestra perspectiva como agentes que tratan de preservarse a sí mismos. Después de todo, una avalancha puede destruir un pueblo y matar gente

del mismo modo que puede hacerlo un ejército entregado al pillaje, e incluso los sencillos átomos de helio pueden empujar contra la parte interior de un globo y mantenerlo tenso. Las enzimas pueden ser unos pequeños agentes ciertamente activos. En realidad, sospecho que es nuestra *incapacidad* para encajar fácilmente los eventos subatómicos en los conceptos de agencia a los que estamos acostumbrados lo que convierte el mundo de la física cuántica en un campo tan extraño y difícil de concebir. Tal como veremos en el siguiente capítulo, los familiares conceptos de causa y efecto están mucho mejor instalados en nuestro mundo macroscópico de agentes que en el mundo subyacente de la microfísica.

EL NACIMIENTO DE LA EVITABILIDAD

Es hora de recapitular y considerar algunas objeciones que he estado dejando para más adelante. El principal objetivo de este capítulo es mostrar que debemos tomarnos en serio la etimología del término *inevitable.* Significa «imposible de evitar». Curiosamente, su negación no se halla actualmente en uso,[4 y *] aunque no cuesta mucho acuñar el término y observar que para cada agente dado existen algunas cosas que son *evitables* y otras en cambio que no lo son. Hemos visto que en un mundo determinista como el mundo Vida podemos diseñar cosas que están más capacitadas que otras para evitar daños en ese mundo, y esas cosas deben su persistencia misma a esta capacidad. Entre todas las cosas que vemos en un plano particular de Vida, ¿cuáles seguirán allí dentro de mil millones de unidades temporales? Las mejores opciones están del lado de aquellas que sean más capaces de evitar el daño. Podemos formular la tesis central del capítulo como la conclusión de un argumento explícito:

> **En algunos mundos deterministas hay entes capaces de evitar daños.**
> **Por lo tanto, en algunos mundos deterministas algunas cosas son evitadas.**
> **Todo lo que es evitado es evitable.**
> **Por lo tanto, en algunos mundos deterministas no todo es inevitable.**
> **Por lo tanto, el determinismo no implica la inevitabilidad.**

4. El *Oxford English Dictionary* incluye *evitable* como un término recogido por primera vez en 1502, actualmente obsoleto y usado sólo en su forma negativa.

*** En este pasaje el autor contrapone dos términos distintos en inglés: *inevitable* y *unavoidable,* ambos traducibles en castellano por «inevitable», como también sus versiones afirmativas, por lo que sus disquisiciones en este punto no se aplican a nuestro idioma. (N. *del t.)***

Este argumento resulta algo sospechoso, ¿no es verdad? Eso es porque pone de manifiesto algunas presunciones ocultas en relación con la evitación y la inevitabilidad que habitualmente no se tienen en cuenta. Resulta extraño señalar casos particulares de evitación como prueba de una «evitabilidad» porque va en contra de una forma muy extendida de pensar sobre la inevitabilidad:

> **Si el determinismo es verdadero, entonces todo cuanto ocurre es el resultado *inevitable* del conjunto de las causas que operan en cada momento.**

Es posible que esta manera de hablar nos resulte familiar, pero ¿qué significa exactamente? Compárese con la siguiente proposición verdadera, aunque trivial:

> **Si el determinismo es verdadero, entonces todo cuanto ocurre viene *determinado* por el conjunto de las causas que operan en cada momento.**

Si «inevitable» no es un sinónimo de «determinado», ¿qué añade a este concepto? ¿Resultado inevitable? ¿Inevitable por quién? ¿Inevitable por el universo en conjunto? Eso no tiene sentido, puesto que el universo no es un agente que tenga interés en evitar nada. ¿Inevitable por cualquiera? Pero eso es falso; acabamos de aprender a distinguir los entes hábiles en la evitación de daños de sus parientes menos talentosos en algunos mundos deterministas. Cuando decimos que cierto resultado particular es inevitable, tal vez nos estemos refiriendo a que es inevitable para todos los agentes presentes en aquel momento y en aquel lugar, pero si eso es cierto o no es independiente del determinismo. Depende de las circunstancias. Todo esto precisa mayor desarrollo, y quién mejor para ayudarme a desarrollarlo que Conrad, el defensor del lector:[5]

5. Conrad es el primo de Otto, el personaje ficticio encargado de articular varias objeciones y críticas a mi teoría de la conciencia en *La contienda explicada*. En diversas reseñas Otto ha sido descrito como mi «marioneta» o mi «conciencia», pero para bien o para mal lo que hacía era expresar de la forma más vivida y convincente que fui capaz de conseguir las dudas más comunes que suscitaban mis ideas en relación con aquel tema. Todo cuanto dice Conrad en este libro es el resultado de la destilación y —en la medida de mis capacidades— el refinamiento de las dudas que han suscitado las tesis del libro. A menudo habla por los críticos a los que doy las gracias en el Prefacio y, si todo sale como espero, también descubrirá que a menudo habla por usted.

CONRAD: **Las configuraciones del mundo Vida que evitan casualmente —o aparentemente— tal o cual cosa no están evitando nada *en realidad*. En último término, todos «viven» en un mundo determinista, y si volvemos a pasar la cinta un millón de veces, volverán a «hacer» exactamente lo mismo —ocurrirá exactamente lo mismo— con independencia de cuánta «evolución» haya habido en aquel mundo. En el escenario evolutivo del mundo Vida, cada evitador en particular, de acuerdo con la localización exacta que ocupa en el plano, cumple con el destino particular que se le había asignado desde el principio, tanto si evita el daño hasta poder replicarse como si no lo consigue. Si supera mil pruebas de «evitación» antes de ser eliminado, ésa es exactamente la vida que le correspondía vivir. Antes hablaba usted de evitadores con «mejores opciones» de sobrevivir, pero evidentemente nadie tiene ninguna opción aquí. Los que sobreviven, sobreviven, y los que no, no, y todo ello está determinado desde el principio.**

Tal como veremos en el próximo capítulo, hay un concepto perfectamente válido de *opción* que es compatible con el determinismo, y que es además el concepto que invocamos para explicar la evolución, entre otras cosas. (La evolución no depende del indeterminismo.) Pero, mientras tanto, tiene usted razón en que cada trayectoria en el mundo Vida está perfectamente determinada, pero ¿por qué insiste en decir que una evitación determinada no es una verdadera evitación? El amplio proceso del que este sencillo evitador (o pseudoevitador, si insiste en llamarlo así) forma parte de manera inconsciente, sólo porque así se han dado las cosas y cumpliendo con su «destino» particular, tiene un poder considerable: produce gradualmente (pseudo)evitadores cada vez mejores, cada vez más capaces de enfrentarse a los problemas de Vida, aunque, por supuesto, tales problemas se hacen también cada vez más severos; es una competición cerrada. El hecho de que el conjunto del proceso esté determinado no quita nada al hecho de que a medida que pasa el tiempo va surgiendo algo que tiene cada vez más el aspecto de evitación.

CONRAD: **Tal vez tenga aspecto de evitación, pero no es *verdadera* evitación. La verdadera evitación consiste en hacer que no ocurra algo que efectivamente iba a ocurrir.**

Supongo que todo depende de lo que quiera usted decir con «iba a ocurrir». Tal vez le confunda la simplicidad de los ejemplos imaginarios del mundo Vida. Existe efectivamente una diferencia entre las respuestas meramente «automáticas» y otras variedades más sofisticadas de evita-

ción, pero no puede usarla para contraponer la evitación en el mundo real con la evitación en el mundo Vida. Un buen ejemplo es el reflejo del parpadeo, que se dispara con extrema facilidad en nosotros, de modo que la mayoría de las veces, cuando parpadeamos en respuesta a algo que surge por sorpresa, se trata de una falsa alarma. En realidad no había ninguna partícula destinada a nuestros ojos; no había nada frente a lo que nuestros párpados hubieran de formar una barrera temporal. Al hacer el balance entre el gasto energético y la breve interrupción de la visión que supone la acción, y los costes de dejar pasar una oportunidad de parpadear que tal vez podría salvar un ojo, la Madre Naturaleza ha «optado por la precaución», probablemente porque los costes (en tiempo y energía) de conseguir más información antes de actuar se disparan muy rápidamente. Los parpadeos son, en general, *involuntarios*, pero hay otras reacciones que sí podemos controlar. El cerebro humano dedica un elaborado subsistema a analizar el movimiento,

> **aunque la mayor parte del espacio de representación está consagrado al cono de direcciones que interseccionan con la cabeza. De nuevo, el sentido de este modelo representacional es intuitivamente evidente: nuestro mayor «interés» se dirige a los objetos que se acercan rápidamente a nuestra cabeza. Es decir, intuitivamente, lo que nos interesa es la pelota de béisbol que nos va a dar directamente en la cara, no la pelota que se dispone a pasar por encima de nuestra espalda izquierda, hecho que queda perfectamente reflejado en nuestro sistema representacional (Akins, 2002, pág. 233).**

Pero, ¿en qué sentido puede decirse que esa pelota «iba a» darnos en la cara? La esquivamos; nos vimos *impulsados* a esquivarla por el elaborado sistema que la evolución ha construido en nosotros para responder a los fotones que rebotan contra los misiles que se acercan en ciertas trayectorias. Nunca «iba a» darte realmente, precisamente porque puso en marcha tu sistema de evitación. Pero ese sistema de evitación es más sofisticado que el simple reflejo del parpadeo, y puede responder a informaciones ulteriores, cuando estén disponibles, y dictar una contraorden a la decisión inicial. Viendo que podemos ganar el partido para nuestro equipo si recibimos el golpe, podemos decidir no esquivarlo. Evitamos evitar algo aunque teníamos la posibilidad de hacerlo, gracias al (*impulsados* por el) conocimiento previo que teníamos del contexto en sentido amplio. Y también podemos evitar evitar la evitación, cuando las circunstancias lo justifiquen. Esta capacidad abierta de los seres humanos está lejos de las

sencillas configuraciones dirigidas a eludir el daño que hemos imaginado en el mundo Vida, pero se equivoca usted si piensa que en el mundo Vida sólo pueden evolucionar reflejos sencillos, «automáticos». Todos los niveles de sensibilidad y reflexión que poseemos los seres humanos son accesibles en principio a las configuraciones de Vida. Al fin y al cabo, en el mundo Vida hay Máquinas Universales de Turing.

CONRAD: **Ya veo adonde quiere ir a parar, pero sigo pensando que lo que ocurre en el mundo Vida, sea cual sea su complejidad o su sofisticación, no puede considerarse una genuina evitación, pues ésta supone** *cambiar el resultado.* **La evitación determinada no es una verdadera evitación porque no cambia el resultado de forma efectiva.**

¿Cambiar qué por qué? La idea misma de cambiar un resultado, por muy corriente que sea, es incoherente, a menos que signifique cambiar el resultado *previsto*, lo que tal como hemos visto es exactamente lo que ocurre en la evitación determinada. El resultado *verdadero,* el resultado *efectivo,* es aquel que se produce, y nada puede cambiarlo en un mundo determinado... ¡ni en uno indeterminado!

CONRAD: **Sin embargo, todas esas entidades del mundo Vida que están en posesión de esos variados poderes para una supuesta evitación, tienen** *inevitablemente* **los poderes que tienen y están** *inevitablemente* **allí donde están en el mundo, en todos y cada uno de los momentos, como resultado de la combinación de las reglas deterministas de dicho mundo y de la posición inicial de la que parte.**

No; éste es precisamente el uso de «inevitablemente» que pretendo cuestionar. Si todo cuanto usted quiere decir es que las capacidades que tiene cada uno de ellos para evitar cosas están *determinadas por el pasado,* entonces tiene razón, pero debe romper con esta mala costumbre suya de asociar el determinismo con la inevitabilidad. *Esa* es la idea con la que debe romperse desde el principio, pues si no es aplicable al acto de esquivar —o no esquivar— la pelota de béisbol, tampoco será aplicable a las muchas proezas palpables de evitación exhibidas por evitadores más simples en el mundo determinista de Vida. Si queremos comprender el mundo biológico, necesitamos un concepto de evitación que pueda aplicarse sin restricciones a los eventos de la historia de la vida en la Tierra, sea o no una historia determinada. Este es, propongo yo, el concepto *propio* de evitación, una evitación tan real como pueda haberla.

Vale la pena señalar, por último, que del mismo modo que la evitabilidad es compatible con el determinismo, la inevitabilidad es compatible con el indeterminismo. Algo es inevitable *para nosotros* si no podemos hacer nada por evitarlo. Si nos mata un rayo indeterminado, podemos decir, retrospectivamente, que no había nada que pudiéramos hacer para evitarlo. No disponíamos de ningún aviso previo. En realidad, si nos enfrentáramos a la perspectiva de correr por un descampado donde los rayos fueran a ser un problema, sería mucho mejor para nosotros que su ritmo y su localización viniera determinada por algo, pues entonces *podrían* ser predecibles para nosotros y, por ello mismo, evitables. El determinismo es el aliado, no el enemigo, de aquellos a quienes no les gusta la inevitabilidad.

Esto debería bastar para romper el vínculo tradicional, o tal vez usual, que se establece entre el determinismo y la falta de esperanza. Hay otros hábitos de pensamiento familiares con los que también deberíamos romper, o a los que al menos deberíamos dedicar una mirada escéptica. Hablar de prevención o evitación en el universo prebiológico o abiológico supone proyectar un concepto más allá de su terreno propio, que parte de nuestra propia imagen como agentes; tal vez no sea una proyección ilusoria en todos los casos, pero siempre implica un riesgo de caer en implicaciones no deseadas. ¿Cuánta prevención hay en nuestro mundo? Decimos que la gravedad impide que un cohete falto de propulsión llegue a ponerse en órbita, porque éste es un tema que nos interesa. Es menos probable que digamos que la gravedad impide que la cerveza del vaso salga flotando por la habitación, pero no porque sea una regularidad menos fiable. Mientras usted lee esto, el latido de su corazón pospone su muerte, y la atención que dedica a la página le impide ver toda clase de otras cosas que hay en su entorno inmediato. Es posible que esté evitando un esguince de tobillo por el hecho de no caminar en este momento, aunque también está acelerando el deterioro de la silla en la que está sentado. No cuesta mucho proyectar escenarios en los que dichas regularidades aparecen dramáticamente representadas como casos de prevención, habilitación, frustración, desvío, enmienda, compensación y otros por el estilo, lo cual es a menudo una perspectiva útil sobre estas regularidades; sin embargo, deberíamos reconocer este hábito o estrategia de pensamiento como lo que es, es decir, una proyección antropocéntrica (o al menos agentocéntrica).

CONRAD: **De acuerdo. Ya veo que no puedo utilizar el término «inevitable» en el sentido habitual, pero sigo teniendo una fuerte sospecha de que me está haciendo *algún* tipo de trampa. Pienso que debe haber algún sentido**

del término «inevitable» de acuerdo con el cual lo que ocurre en un mundo determinado sea inevitable. Y no veo nada de lo que *yo* llamo libertad en el mundo Vida.

De acuerdo. En los próximos capítulos seguiremos atentos a este elusivo sentido de «inevitable», pero mientras tanto deberá usted conceder que he invertido la carga de prueba: ya no cabrá inferir *en ningún sentido* la inevitabilidad del determinismo a menos que se aporten argumentos suplementarios. Y estoy de acuerdo en que nos encontramos aún muy lejos del libre albedrío. No hay nada que se parezca remotamente a la libertad en el nivel de la física del mundo Vida. Los planeadores y los comilones no son libres en absoluto, y hacen simplemente lo que tienen que hacer, siempre y en todo momento. Parece conforme a razón decir que nada que esté compuesto por dichas partes carentes de libertad puede disfrutar de ninguna libertad superior a ellas, que *el conjunto no puede ser más libre que sus partes,* pero esta intuición, que se encuentra en la base misma de la resistencia al determinismo, se revelará más adelante, tras un examen detallado, como una ilusión. En el próximo capítulo examinaremos más de cerca esta concepción basada en la perspectiva del agente sobre la causa y el efecto, la posibilidad y la oportunidad, para ver con más detalle por qué la importante cuestión de la inevitabilidad no tiene nada que ver con la cuestión del determinismo.

Capítulo 2

Una versión modelo del determinismo demuestra que en el Vasto espacio de las configuraciones posibles de la «materia» hay algunas que persisten mejor que otras, porque han sido diseñadas para evitar el daño. El proceso por el que emergen estas entidades emplea información obtenida de su entorno para anticipar rasgos generales y a veces particulares de futuros probables, lo que permite una acción guiada por dicha información. Esto demuestra que es posible conseguir la evitabilidad *en un mundo determinista y, por lo tanto, que la asociación corriente entre el determinismo y la inevitabilidad es errónea. El concepto de inevitabilidad, igual que el concepto de evitación en el que se basa, pertenecen propiamente al nivel del diseño, no al nivel físico.*

Capítulo 3

Los conceptos de causa y posibilidad se hallan en la base de la inquietud acerca de la libertad, y un análisis demuestra que nuestros conceptos ordinarios no tienen las implicaciones que a menudo se les atribuyen: el determinismo no es ninguna amenaza para nuestras ideas centrales respecto al papel que desempeñan las posibilidades y las causas en nuestras vidas.

NOTAS SOBRE FUENTES Y LECTURAS COMPLEMENTARIAS

Pueden encontrarse argumentaciones más extensas para defender las conclusiones de este capítulo en «Real Patterns» (1991b), *La peligrosa idea de Darwin* (1995), *Tipos de mente* (1996a) y, más recientemente, «Collision Detection, Muselot, ant Scribble: Some Reflections on Creativity» (2001a).

Paul Rendell ha creado una máquina de Turing «simple» en el mundo Vida, la cual es (en principio, no en la práctica) expandible hasta obtener una Máquina Universal de Turing; se la puede ver y explorar en su página web: http://www.rendell.uk.co/gol/tm.htm. La lista de sus partes —todas diseñadas a partir de planeadores, comilones y otros miembros de la familia— resulta inspiradora: lEspacio3, lEspacio4, lEspacio8, Dirección de Columnas, Comparador, Control de Conversión, Distribuidor, Máquina de Estado Finito, Puerta de Entrada, Célula de Memoria, Metamorfosis II, Cañón MWSS, Retraso de Estado Siguiente, Puerta NOT XOR, Puerta de Salida, Colector de Salida, Cañón P120, Cañón P240, Cañón P30LWSS, Control de Pop, Control de Empuje, Dirección de Remo, Pestillo de Reinicio (a), Pestillo de Reinicio (b), Detector de Señal, Pila, Célula de Pila, Comida para llevar, Cinta de Turing.

Capítulo 3

Pensar el determinismo

El determinismo parece robarnos nuestras oportunidades, sellar nuestros destinos en una red global de cadenas causales que se remontan indefectiblemente al pasado. Todos nosotros pasamos buena parte de nuestro tiempo pensando cómo *podrían* irnos las cosas hoy o el año que viene, o cómo *podrían* habernos ido si hubiera pasado esto o lo otro. En otras palabras, parecemos presuponer que nuestro mundo no es determinista.

MUNDOS POSIBLES

En nuestras deliberaciones distinguimos sin problemas entre cosas que podrían haber ocurrido y cosas que no podrían haber ocurrido, cosas que no ocurrirán en ningún caso y cosas que podrían ocurrir perfectamente, si así lo queremos. Tal como dicen los filósofos, con frecuencia imaginamos *mundos posibles*'.

> En el Mundo A, los disparos de Oswald no hicieron blanco en Kennedy sino en LBJ, lo que introdujo millones de cambios en la historia subsiguiente.

Y usamos estos productos de nuestra imaginación para guiar nuestras elecciones, aunque sólo un filósofo utilizaría fórmulas como las siguientes para expresarlo:

> Acabo de imaginar un mundo exactamente igual al presente excepto porque no he comido ese pastel de crema y, por lo tanto, no experimento el remordimiento que siento ahora.
>
> En el Mundo A, le propongo matrimonio a Rosemary. En el Mundo B, le envío esta nota de despedida que estoy escribiendo e ingreso en una orden monástica.

Por muy familiar que nos resulte este ejercicio de imaginación, a menudo nos juega malas pasadas cuando tratamos de pensar rigurosamente el determinismo y la causalidad. En el presente capítulo sostendré que el determinismo es enteramente compatible con las presuposiciones que gobiernan nuestra manera de pensar acerca de lo que es posible. La aparente incompatibilidad es, lisa y llanamente, una ilusión cognitiva. Tal conflicto no existe. Tanto en nuestras reflexiones cotidianas sobre lo que vamos a hacer a continuación como en nuestro pensamiento científico más riguroso en relación con las causas de los fenómenos, empleamos conceptos de *necesidad, posibilidad* y *causalidad* que son estrictamente neutrales respecto a la cuestión de si la verdad está del lado del determinismo o del indeterminismo. Si estoy en lo cierto, más de un eminente filósofo está equivocado, por lo que cabe esperar bastante artillería pesada (aunque no será más que un rumor en la distancia, porque no pienso presentarles batalla en este libro). Christopher Taylor me ha ayudado mucho a clarificar mi pensamiento en relación con este tema y me ha mostrado el modo de lanzar una campaña más profunda y radical en apoyo de mis tesis iniciales, y el artículo que hemos escrito en colaboración (Taylor y Dennett, 2001) aporta mayores detalles técnicos de los necesarios aquí. Voy a presentar una versión más asequible del argumento, y subrayaré las ideas principales para que los no-filósofos puedan al menos ver dónde están los puntos en disputa y cómo proponemos resolverlos, y dejaré a un lado casi todas las fórmulas lógicas. Los filósofos pueden consultar la versión completa, por supuesto, para ver si hemos atado realmente los cabos sueltos y cubierto las lagunas que pasamos por alto en esta exposición. Y como lo que sigue se debe en gran medida a Taylor, habrá un cambio temporal de los pronombres de autor al «nosotros».

Nuestra tarea consiste, pues, en clarificar los conceptos *cotidianos* de posibilidad, necesidad y causalidad que forman parte de nuestras reflexiones, proyectos, preocupaciones y fantasías, y que nos ayudan a enfrentarnos al mundo y a sus desafíos. Podemos hacernos la tarea más fácil si restringimos nuestras reflexiones sobre los mundos posibles a los universos democriteanos de Quine. Es conocido el escepticismo de Quine hacia cualquier intento de hablar seriamente sobre la posibilidad y la necesidad —el tema de la *lógica modal*—, y sus universos democriteanos están diseñados para proporcionar la base de operaciones más manejable y ordenada posible para explorar estas cuestiones. Como se recordará del capítulo 2, el Vasto número de universos democriteanos consiste en una colección de puntos-átomos cuyas trayectorias a través del espacio y el tiempo vienen dadas por sus coordenadas tetradimensionales $\{x, y, z, t\}$. Una com-

pleta *descripción de estado* del mundo en un tiempo *t* es simplemente el catálogo exhaustivo de las direcciones ocupadas {*x, y, z,*} en *t.* Llamamos al conjunto de todos los mundos *lógicamente* posibles la Biblioteca de Demócrito, y podemos llamar al subconjunto que contiene sólo los mundos *físicamente posibles* Φ (phi). Por supuesto, no conocemos aún todas sus leyes físicas, y no podemos saber con certeza si son deterministas o indeterministas, pero podemos pretender que los conocemos. (Ahora que disponemos del mundo Vida de Conway, siempre podemos contrastar nuestras intuiciones dentro del mundo Vida de Conway, donde *sí* conocemos perfectamente las leyes físicas y del que sabemos que es determinista.)

Dado un mundo posible, tenemos muchas maneras de realizar aserciones sobre él. Tal como vimos en el caso del mundo simplificado de Vida, lo más natural será saltar por encima del nivel atómico y describir el mundo en términos de pedazos más grandes. Del mismo modo que podemos seguir los pasos de un planeador particular desde su nacimiento hasta su muerte sobre el plano de Vida, podemos seguir el rastro también de las trayectorias en el espacio y el tiempo de «hipersólidos compactos» (objetos tetradimensionales) como las estrellas, los planetas, los seres vivos y demás parafernalia cotidiana (objetos familiares que aparecen en la vida humana). Una famosa imagen de Platón dice que debemos cortar la naturaleza siguiendo sus propias articulaciones, y las articulaciones por las que *nosotros* comenzamos —literalmente, donde una *cosa* termina y comienza la siguiente— son aquellos rasgos que resultan lo bastante prominentes y estables *para nosotros* como para que podamos identificarlos (y seguirles la pista, y volverlos a reconocer) como objetos macroscópicos. Tal como vimos en el mundo Vida, la «física» subyacente (la regla de transición entre estados) dicta qué configuraciones son lo bastante resistentes a lo largo del tiempo como para constituir regularidades macroscópicas (no microscópicas), las cuales sirven de punto de apoyo para nuestra imaginación a la hora de pensar en causas y posibilidades. Podemos describir tales configuraciones de tamaño medio por el método familiar de los *predicados informales* aplicables a dichas entidades, tales como (en orden de menos a más dudosos) «tiene una longitud de un metro», «es rojo», «es humano», «cree que la nieve es blanca». Estos predicados informales plantean una infinidad de problemas relacionados con la vaguedad, la subjetividad y la intencionalidad, y son estos problemas —los problemas que surgen cuando saltamos del nivel básico de los átomos y el espacio a categorías ontológicas superiores— los que alimentaron el escepticismo de Quine acerca de la posibilidad de hablar con sentido sobre la posibilidad y la necesidad. Pensamos

que si hacemos explícito este paso y si concentramos todos los puntos oscuros en el desplazamiento del nivel físico atómico al nivel cotidiano, podemos mantener dichos problemas aislados y evitar que pongan en peligro nuestro planteamiento básico. Si procedemos, pues, con cautela y suponemos que podemos apoyarnos tentativamente en los predicados informales, no nos creará muchos problemas de conciencia formar proposiciones como

(1) Hay algo que es humano

y determinar si son aplicables a diferentes mundos posibles. No hay seres humanos en ningún mundo Vida, puesto que los seres humanos son seres tridimensionales, pero en algunos de ellos podría haber entidades bidimensionales prodigiosamente parecidas a los seres humanos. Llevando la cuestión a un terreno más cercano: ¿cabría considerar que hay algo humano en un mundo posible donde las criaturas bípedas parlantes, tecnológicas y culturales tuvieran plumas en lugar de pelo en la cabeza y descendieran de las avestruces? ¿O tal vez consideraríamos que tal criatura es una *persona* no humana? ¿Es «humano» una categoría biológica o, tal como sugiere la palabra «humanidades», una categoría sociocultural y política? Es posible que haya opiniones divergentes respecto a la interpretación del predicado informal «humano». A menudo encontraremos términos fronterizos sobre los que nos resultará difícil llegar a veredictos incontestables.

Merecen una atención especial los *predicados de identificación* de la forma «es Sócrates». «Es Sócrates», cabe suponer, es aplicable a cualquier entidad de cualquier mundo posible que comparta suficientes rasgos con el conocido habitante de nuestro mundo como para que estemos dispuestos a considerarlo la misma persona. En nuestro mundo, por supuesto, «es Sócrates» se aplica a una única entidad; en otros mundos, puede ser que no haya tal ser, o que haya uno, o incluso cabe concebir que haya dos o más seres a quienes el predicado se aplique con la misma validez. Igual que sucede con otros predicados informales, los predicados de identificación adolecen de vaguedad y subjetividad, pero es posible aislar y manejar estas fastidiosas cuestiones cuando se plantean en casos particulares.[1]

1. Aviso para expertos: en efecto, estamos pasando por alto los grandes debates en relación con la designación rígida, y asumimos el riesgo. Que nos pillen si pueden. (La designación rígida es un concepto que debemos a Kripke [1972], y hay opiniones divididas sobre si consigue realmente resucitar el esencialismo. Nosotros pensamos que no, pero preferiríamos no dedicar el resto del año a defender nuestra posición.)

Ahora ya estamos en condiciones de definir los conceptos fundamentales que precisamos —*necesidad, posibilidad* y *causalidad*— en términos de mundos posibles. Una proposición del tipo

(2) Necesariamente, Sócrates es mortal

puede traducirse como

(3) En todo mundo (¿físicamente?) posible /, la proposición «Si algo es Sócrates, entonces es mortal*» es verdadera.

En otras palabras, cuando nos ponemos a examinar todas las posibilidades que somos capaces de contemplar, descubrimos que no hay ni un solo mundo posible en el que haya un Sócrates inmortal. *Eso es lo que significa* decir que Sócrates es necesariamente mortal. En este caso «es Sócrates» y «es mortal» son predicados informales del tipo que acabamos de ver. Sin duda, decidir si la proposición es verdadera plantea muchos problemas, que proceden en gran medida del carácter inevitablemente ambiguo de los predicados: ¿es un candidato a Sócrates que sea mortal pero que pueda volar como Superman menos merecedor del predicado «es Sócrates» que un candidato a Sócrates incapaz de elevarse del suelo pero que milagrosamente no se vea afectado por su copa de cicuta? ¿Quién puede resolver la cuestión? Es más, todavía no hemos decidido si el conjunto de mundos posibles entre los que podría estar /debería ser la Biblioteca de Demócrito al completo (todos los mundos), o í> (los mundos físicamente posibles), o incluso algún conjunto X más restringido. La lógica sola no puede resolver esta cuestión, pero el lenguaje lógico sí nos ayuda a poner de relieve tales cuestiones y señalar con mayor precisión el tipo de vaguedad al que nos enfrentamos.

Ahora podemos definir la *posibilidad.* Lo *posible* es todo aquello que no es *necesariamente falso,* de modo que

(4) Es posible que Sócrates hubiera tenido el pelo rojo

significa

(5) Existe (al menos un) mundo posible /en el que la proposición «Hay algo que es Sócrates y que tiene el pelo rojo» es verdadera.

Una vez más, debemos decidir si estamos hablando de posibilidad física o lógica. Tal hipótesis es *físicamente* posible si existe un mundo dentro del conjunto O donde hay un Sócrates pelirrojo. En caso contrario, queda descartado desde el punto de vista físico, por más comunes que sean los Sócrates pelirrojos en los mundos lógicamente posibles pero físicamente imposibles.

Ahora ya estamos en condiciones de clarificar la definición del determinismo ofrecida al comienzo del capítulo 2: *en cada momento dado hay exactamente un único futuro físicamente posible.* Decir que el determinismo es verdadero es decir que nuestro mundo forma parte de un subconjunto de mundos que poseen la interesante propiedad siguiente: no hay dos mundos que comiencen exactamente igual (si comienzan siendo iguales, siguen siendo iguales para siempre: no son mundos *diferentes* en ningún sentido), y siempre que dos mundos compartan *cualquiera* de sus descripciones de estado, compartirán también todas las descripciones de estado posteriores a ésta. El mundo Vida ilustra claramente esta idea. Sólo es determinista en una dirección; en general no se puede extrapolar el intante *previo,* mientras que siempre se puede extrapolar el instante siguiente. Por ejemplo, un plano Vida que contenga una naturaleza muerta cuadrada de cuatro píxels en el tiempo *t* (véase la figura 3.1) tiene un pasado ambiguo. El próximo estado (y el siguiente, y el siguiente) es exactamente el mismo —a menos que irrumpa algo—, pero el estado previo podría haber sido cualquiera de estos cinco (o una cantidad indefinida de configuraciones con píxels que se apagan a mayor distancia).

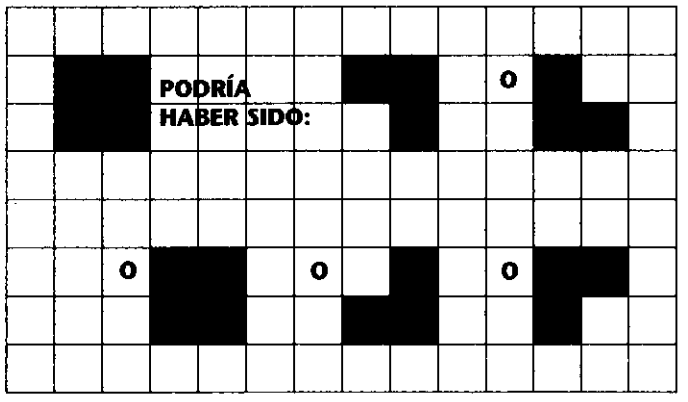

FIGURA 3.1. Naturaleza muerta y lo que podría haber sido.

En consecuencia, si el deterninismo —así definido— es verdadero, podemos concluir que incluso aunque muchos pasados, diferentes pudieran haber llevado exactamente al estado presente, nuestro futuro está «fijado» por nuestro estado presente. Desde esta perspectiva, el determinismo *parece* ser exactamente lo opuesto a nuestro punto de vista corriente, según el cual el pasado está «fijado» y el futuro está «abierto.» Podríamos definir una versión más fuerte (y no estándar) del determinismo que excluya tal ambigüedad en el pasado, y descarte lo que llamo *hechos históricos inertes*, es decir, hechos sobre el pasado que, por lo que respecta a nuestras leyes de la física, podrían haber sido de una manera o de otra sin dejar ningún efecto subsiguiente. La capacidad de los cosmólogos de «rebobinar la película» y calcular de este modo los hechos relativos a los primeros momentos posteriores al Big Bang muestra que, en relación con ciertas propiedades, podemos leer el pasado a partir del presente con sorprendente precisión y alcance, pero esto no sirve en absoluto para demostrar que no haya hechos históricos inertes. El hecho de que parte del oro que tengo en los dientes perteneciera una vez a Julio César —o su negación, el hecho de que no le perteneció jamás— es un ejemplo plausible de un hecho histórico inerte. No hay duda de que es *prácticamente* inerte. Como se da el caso de que no mantenemos ningún registro de la cadena de propietarios de los fragmentos de oro del mismo modo que lo hacemos, por ejemplo, con las pinturas de Rembrandt, es prácticamente inimaginable que ninguna investigación del estado presente de la distribución de átomos del mundo pudiera permitir descubrir cuál de esas dos proposiciones es la verdadera, aunque seguramente una de ellas lo es.

Y cuando miramos hacia el futuro, resulta casi imposible decir cuándo un hecho histórico inerte *hasta el momento* vendrá a «marcar una diferencia» en lo que pase a continuación. Supongamos que el determinismo es verdadero y que conocemos a la perfección las leyes de la física, tal como las conoce el demonio de Laplace. Sin embargo, a menos que tengamos un conocimiento perfecto y completo de una descripción de estado del universo, no seremos capaces de decir cuál es el mundo en el que vivimos dentro del Vasto número de universos posibles con diferencias microscópicas que forman parte del conjunto O. Es precisamente porque nuestro conocimiento es inevitablemente incompleto por lo que es un recurso tan bueno pensar en términos de mundos posibles.

Una de las aplicaciones más útiles del modelo de los mundos posibles consiste en usarlo para la interpretación de proposiciones hipotéticas, como

(6) Si Greenspan hubiera llorado en el Congreso, el mercado se habría desplomado

y

(7) Si le hubiéramos puesto la zancadilla a Arthur, habría caído.

Siguiendo a David Lewis (1973), vemos que la proposición (7) es *(aproximadamente)* verdadera, y que si y sólo si en cualquier mundo *aproximadamente parecido* al nuestro se cumpliera el *antecedente*, lo mismo sucedería con el *consecuente*. En otras palabras,

(8) Tomemos el conjunto X de mundos parecidos a nuestro mundo presente: en todos los mundos de este conjunto se verifica que si *le ponemos la zancadilla a Arthur*, también se verifica que *Arthur cae*.

A veces, cuando hacemos afirmaciones hipotéticas de este tipo, tenemos tendencia a comprobarlas por el procedimiento de *imaginar* unas cuantas variaciones. («Veamos: supongamos que Arthur llevara una camisa roja, ¿le hubiera impedido *eso* caer? Supongamos que la radio estuviera apagada, supongamos que la calefacción estuviera encendida, supongamos que llevara rodilleras... No, hubiera caído igualmente. Supongamos que la habitación estuviera llena de bolsas de aire hinchadas o que el edificio entero estuviera en caída libre a gravedad cero; *eso* hubiera impedido que cayera... Pero eso es demasiado distinto de nuestro mundo como para que pueda valer.») Y en experimentos controlados, no nos limitamos sólo a imaginar, sino que investigamos realmente estas variaciones. Alteramos metódicamente las condiciones y vemos cuáles son los cambios que se verifican. Tal como veremos, esto no es un procedimiento tan evidente como podría parecer a primera vista.

Con independencia de si realizamos algún experimento práctico o mental, la presunción que hay detrás de la aserción hipotética es que hay un conjunto de mundos X parecidos al nuestro donde se cumple esta regularidad. Y, en general, podemos expresar la interpretación de un hecho hipotético como (6) o (7) del siguiente modo:

(9) En el conjunto de mundos X, A => C,

donde A es el antecedente y C el consecuente.

Pero *¿cuan parecidos* a nuestro mundo deben ser los mundos del conjunto X? No siempre es fácil en estos casos escoger un valor óptimo para X, pero podemos seguir las siguientes directrices:

En proposiciones como (6) y (7), X debería:

- contener mundos en los que A se cumple, no-A se cumple, C se cumple, y no-C se cumple,
- contener mundos parecidos al mundo presente en todo lo demás (en la medida en que lo permita la cláusula precedente).

De modo que para analizar (7), debemos escoger un conjunto X que contenga mundos en los cuales le ponemos la zancadilla a Arthur, mundos en los cuales nos abstenemos de hacerlo, mundos en los que Arthur cae, y mundos en los que sigue de pie. (Nótese que para reunir estos mundos empleamos nuestras ontologías de nivel superior. *No* graduamos el parecido de los mundos calculando cuántos vóxels están llenos de hierro o de oro; para determinar qué mundos queremos incluir usamos los predicados informales, a pesar de la imprecisión y la vaguedad que comportan. Tal como veremos, se da el caso que muchos de los dilemas que surgen a propósito de la causalidad y la posibilidad dependen de cómo escojamos definir X, el conjunto de referencia de mundos posibles vecinos.)

CAUSALIDAD

Por último, ¿qué pasa con la *causalidad*? Algunos filósofos esperan descubrir algún día una descripción «verdadera» de la causalidad, pero dada la naturaleza informal, vaga y a menudo contradictoria del término, pensamos que un objetivo más realista sería desarrollar un análogo (o análogos) formal que nos ayudara a pensar más claramente en el mundo. Nuestras intuiciones previas sobre la causalidad nos servirán de guía, pero deberíamos desconfiar siempre de que alguien pretenda hacer pasar un argumento informal por una «prueba» capaz de validar o desacreditar doctrinas particulares sobre la causalidad.[2] Cuando decimos

2. Por supuesto, no dejará de haber filósofos dispuestos a discutir estas palabras. Ningún problema; les traspasamos con el mayor gusto la carga de prueba a ellos. Si pueden presentar una teoría exenta de problemas y contraejemplos sobre el concepto ordinario de causalidad *completo,* compararemos con dicha teoría nuestro proyecto más modesto y esquemático para ver si nos hemos dejado algo importante. En espera de eso, seguiremos adelante con nuestro análisis basado en una versión parcial que incluye lo que nos parecen los aspectos más importantes del concepto cotidiano.

(10) La zancadilla que Bill le puso a Arthur hizo que éste cayera

hay una serie de factores que parecen venir en apoyo de nuestra aserción. En un orden aproximado de importancia, la lista sería la siguiente:

- *Necesidad causal*: nuestro asentimiento a la proposición (10) depende de nuestra convicción de que si Bill no le hubiera puesto la zancadilla a Arthur, éste no hubiera caído. Aplicando la interpretación de las hipótesis que acabamos de proponer, escogemos como conjunto X una colección de mundos parecidos al nuestro que incluye mundos en los que (i) Bill le pone la zancadilla a Arthur; mundos en los que (ii) Bill no le pone la zancadilla; mundos en los que (iii) Arthur cae; y mundos en los que (iv) Arthur no cae. Y nos aseguramos de que en todos los mundos del conjunto X donde Arthur cae, Bill le puso la zancadilla.
- *Suficiencia causal*: es muy posible que cuando afirmamos (10), lo hagamos en parte porque creemos que la caída de Arthur era un *resultado inevitable* de la zancadilla de Arthur: en *todos* los mundos donde Bill pone un obstáculo en el camino de Arthur, éste se da de bruces. (De nuevo encontramos la palabra «inevitable» [*inevitable*], y efectivamente significa inevitable* [*unavoidable*] en este caso: por una razón u otra, Arthur no puede evitar caer, los amigos de Arthur no pueden impedir que caiga, no hay nada más en la situación que pueda interferir en su caída, y así sucesivamente; en esta ocasión la gravedad tampoco se verá desafiada.) Esta segunda condición es enteramente distinta de la primera desde el punto de vista lógico, a pesar de lo cual ambas aparecen habitualmente confundidas en nuestra manera cotidiana de pensar. En realidad, tal como veremos, muchas veces la confusión tiene su origen precisamente en este punto. Más adelante discutiremos más por extenso las relaciones entre estas dos condiciones.
- *Independencia:* partimos del supuesto de que las dos proposiciones A y C son lógicamente independientes entre sí. Dicho en términos de mundos posibles: es preciso que existan mundos, por muy alejados que estén de la realidad, en los que A se cumple pero no C, y viceversa. Es por ello que la proposición «Que Mary can-

• De nuevo se refiere Dennett a la distinción entre *inevitable* y *unavoidable,* que no tiene equivalente en castellano. (N. *del t.)*

te y baile hace que Mary cante y baile» suena decididamente rara. Esta condición también ayuda a descartar «1 + 1 =2 hace que 2 + 2 = *4*».

- *Prioridad temporal:* una forma fiable de distinguir causas de efectos es señalar que las causas ocurren antes. (Aviso para expertos.)
- *Miscelánea de criterios ulteriores:* existen algunas condiciones ulteriores que, aunque sean menos cruciales que las anteriores, pueden contribuir a aumentar nuestra confianza al hacer juicios causales. Por ejemplo, en los ejemplos de manual sobre la causalidad, A describe a menudo las acciones de un agente, y C representa un cambio en el estado de un objeto pasivo (como en «Mary provocó un incendio en la casa»). Por otro lado, nosotros acostumbramos a esperar que los dos participantes entren en contacto físico durante su interacción.

Para comprender mejor estas condiciones, veamos cómo encajan en algunos casos de prueba, algunos de los cuales están tomados de Lewis (2000). Primero consideremos el caso de un tirador experto que apunta hacia una víctima situada a cierta distancia. Supongamos que un examen del historial del tirador muestra que la probabilidad de que acierte en este caso es de 0,1; por si alguien piensa que puede marcar alguna diferencia, podemos imaginar que el resultado viene determinado en parte también por los eventos cuánticos irreduciblemente azarosos que puedan producirse en el aire que los separa, o en el cerebro del tirador. Supongamos que en el caso presente la bala da efectivamente en la víctima y la mata. En tal caso no dudamos en afirmar que las acciones del tirador fueron la causa de la muerte de la víctima, *a pesar de su insuficiencia causal.* De acuerdo con eso, parece que, *al menos en casos de este tipo,* la gente valora la necesidad por encima de la suficiencia al hacer juicios causales.

Sin embargo, la suficiencia sigue conservando cierta relevancia. Supongamos que tanto el rey como el alcalde están interesados en el destino de cierto joven disidente; se da el caso de que ambos dictan órdenes de exilio contra él; el disidente va al exilio. Éste es un caso típico de *sobredeterminación.* Sea Ai «el rey dicta una orden de exilio», y A_2 «el alcalde dicta una orden de exilio», y C «el disidente va al exilio». Dentro de este escenario, ni $A_:$ ni A_2 son individualmente necesarios para C: por ejemplo, si el rey no hubiera dictado ninguna orden, el disidente hubiera ido al exilio por orden del alcalde, y viceversa. Sin embargo, la suficiencia viene

al rescate y nos permite escoger entre los dos. En este caso A_2 no logra pasar la prueba: resulta fácil imaginar un universo donde el alcalde dicta su decreto y, sin embargo, el disidente logra sustraerse a él (sólo hay que cambiar la orden del rey por su perdón). La orden del rey, en cambio, es verdaderamente *efectiva-,* por más pequeños cambios que introduzcamos en el universo (incluidos cambios en las órdenes del alcalde), la orden del rey tiene como resultado el exilio del disidente. De acuerdo con esto, podemos decir que A] es la «causa real» (si sentimos la necesidad de satisfacer nuestro impulso en este sentido).

Por último, consideremos el caso de Billy y Susie. Ambos niños están jugando a tirar piedras a una botella de cristal, y se da el caso de que la piedra de Susie va un poco más rápida, llega antes a la botella y la rompe. La piedra de Billy llega exactamente al sitio donde estaba antes la botella, pero al hacerlo un momento más tarde no encuentra sino cascos rotos por el aire. Si debemos escoger entre A] («Susie lanza la piedra S») y A_2 («Billy lanza la piedra B»), votamos por A, como causa de C («La botella se rompe»), a pesar del hecho de que ninguna de las proposiciones es necesaria (si Susie no hubiera lanzado su piedra, la botella se habría roto igualmente por la acción de Billy, y viceversa) y ambas son suficientes (el tiro de Billy es suficiente para producir la ruptura de la botella, haga lo que haga su compañera de juegos, y lo mismo sucede en el caso de Susie). ¿Por qué? La noción general de la prioridad temporal (introducida antes en relación con la posibilidad de distinguir la causa del efecto) nos parece una condición determinante en este caso. Igual como sucede en las disputas de prioridad en el campo de la ciencia, el arte y el deporte, parecemos conceder un reconocimiento especial al que aporta una innovación *primero,* y como la piedra S llegó a las inmediaciones de la botella antes que la piedra B, le damos el crédito a Susie. Por otro lado, está claro que, aunque la botella se habría roto igualmente sin el lanzamiento de Susie, el suceso de la ruptura hubiera sido significativamente diferente, al ocurrir en un momento posterior y con una piedra diferente que lanzaría cascos rotos en direcciones diferentes. (Nótese que este problema surge precisamente porque hemos dado el salto a la ontología cotidiana de botellas y rupturas, y sus fastidiosas condiciones de identidad. El problema aquí es lo que debe *valer como* el «mismo efecto», no cualquier incertidumbre subyacente en relación con lo ocurrido.)

Podemos diseñar el conjunto X para que refleje este hecho (de acuerdo con las directrices): deberá contener mundos en los que (1) la botella

no se rompe o (2) se rompe de un modo muy parecido a como se rompe en la realidad. Entonces, para todo mundo de X, se verifica que

$$C \Rightarrow A_1$$

siempre que la botella se rompe en X, encontramos que Susie ha lanzado primero su piedra. Por otro lado,

$$C \not\Rightarrow A_2$$

puede ser perfectamente falso en X; X puede contener ciertamente mundos en los que la botella se rompe pero Bill no llega a tirar su piedra. En resumen, si escogemos bien el conjunto X, A_1 es «más necesario» que A_2. La vaguedad de X, aunque a veces resulte fastidiosa, puede servir también para superar algunos puntos muertos.

Eso no significa que los puntos muertos se puedan superar siempre. Debemos contemplar con ecuanimidad la posibilidad de que a veces las circunstancias impidan señalar una única «causa real» para un evento, por más que la busquemos. Un caso paradigmático es el clásico dilema de las escuelas de derecho:

> **Todo el mundo en los puestos adelantados de la Legión Extranjera francesa odia a Fred y quiere verlo muerto. Durante la noche previa a una expedición de Fred por el desierto, Tom le envenena el agua de la cantimplora. Luego Dick, sin conocer la intervención de Tom, vierte el agua (envenenada) y la sustituye por arena. Finalmente, viene Harry y agujerea la cantimplora, de modo que ésta se va vaciando de «agua». Más tarde, Fred se levanta y se pone en marcha, provisto de su cantimplora. No descubrirá que su cantimplora está casi vacía hasta que sea demasiado tarde, pero además encontrará que lo que queda en ella es arena, no agua, ni siquiera agua envenenada. Como consecuencia, Fred muere de sed. ¿Quién fue el causante de su muerte?**[3]

Muchos sentirán la tentación de insistir en que *debe* haber una respuesta para esta pregunta, y para otras parecidas. Ciertamente podemos

3. Versión doblemente elaborada del ejemplo original de McLaughlin (1925), desarrollado luego por Hart y Honoré (1959). La versión de Hart y Honoré tiene un giro menos: «Supongamos que A entra en el desierto. B pone secretamente una dosis fatal de veneno en el barrilete de agua de A. A lleva el barrilete al desierto, donde C lo roba; tanto A como C piensan que contiene agua. A muere de sed. ¿Quién le ha matado?».

ponemos de acuerdo para legislar una respuesta si sentimos la necesidad de hacerlo, y algunas propuestas legislativas serán sin duda más atractivas, más intuitivas, que otras, pero no queda claro que haya ningún hecho —relativo a cómo es el mundo, o a cuál es el auténtico sentido de nuestras palabras, o incluso a cuál *debería* ser el sentido de nuestras palabras— que pueda resolver la cuestión.

EL TIRO AL HOYO DE AUSTIN

Ahora que tenemos una mejor comprensión de los mundos posibles, podemos exponer tres grandes confusiones relativas a la posibilidad y la causalidad que han oscurecido siempre la búsqueda de una teoría sobre la libertad. En primer lugar, está el miedo a que el determinismo limite nuestras posibilidades. Podemos comprender el *aparente* crédito que merece esta idea si consideramos el famoso ejemplo propuesto hace muchos años por John Austin:

> Consideremos el caso de que fallo un golpe muy corto y me enfado conmigo mismo porque podría haberlo embocado. Eso no significa que debería haberlo embocado si lo hubiera intentado: lo intenté, y fallé. Tampoco significa que debería haberlo embocado si las condiciones hubieran sido distintas de las que fueron: por supuesto que podrían haberlo sido, pero estoy hablando de las condiciones exactamente tal como se dieron, y afirmo que podría haberlo embocado. Ahí está el problema. «¿Puedo embocarlo esta vez?» tampoco significa que vaya a embocarlo esta vez si lo intento o si cualquier otra cosa; pues puede ser que lo intente y falle, y sin embargo eso no me convencería de que no podría haberlo hecho; los experimentos posteriores no harán sino confirmar mi creencia de que podría haberlo hecho aquella vez, aunque no lo hiciera (Austin, 1961, pág. 166).

Austin no embocó su golpe. ¿Acaso lo hubiera conseguido, si el determinismo fuera verdadero? Una interpretación del caso en términos de mundos posibles pone al descubierto el paso en falso del razonamiento de Austin. Primero, supongamos que el determinismo es verdadero y que Austin falla; sea H la proposición «Austin emboca el golpe». Debemos escoger ahora el conjunto X de mundos posibles relevantes que debemos sondear para ver si podría haberlo conseguido. Supongamos que escogemos como X el conjunto de mundos físicamente posibles que son *idénticos* al mundo presente en algún tiempo t_0 previo al golpe. Como el deter-

minismo afirma que en cada momento hay exactamente un solo futuro físicamente posible, dicho conjunto de mundos tiene un solo miembro, el mundo presente, el mundo donde Austin falla. De modo que, si escogemos el conjunto X siguiendo este criterio, el resultado que obtenemos es que H no es cierto para ningún mundo de X. Así pues, de acuerdo con esta interpretación, no era posible que Austin embocara el golpe.

Por supuesto, este método para escoger X (llamémoslo el *método estrecho)* es sólo uno de los muchos posibles. Supongamos que admitiéramos también mundos que contuvieran diferencias microscópicas respecto al mundo presente en t_0\ podría muy bien ser que ahora hubiéramos incluido mundos en los que Austin emboca el golpe, aunque sean deterministas. Este es, después de todo, el resultado al que han llegado los recientes trabajos en relación con el caos: muchos fenómenos de interés para nosotros pueden cambiar radicalmente con sólo pequeñas alteraciones en las condiciones iniciales. De modo que la cuestión es: cuando la gente mantiene que hay acontecimientos posibles, ¿están pensando realmente en términos del método estrecho o no?

Supongamos que Austin es del todo incompetente como jugador, y que su compañero en el partido de parejas de hoy se inclina por negar que hubiera podido embocar el golpe. Si dejamos que X abarque demasiado, corremos el riesgo de incluir mundos en los que Austin, gracias a años de caras lecciones de golf, termina por convertirse en un jugador de campeonato que emboca el golpe con facilidad. Pero cabe presumir que Austin no se refiere a eso. Austin parece suscribir el método estrecho para la elección de X cuando insiste en que está «hablando de las condiciones exactamente tal como se dieron». Y, sin embargo, en la proposición posterior parece echarse atrás de tal suscripción al observar que «los experimentos posteriores no harán sino confirmar mi creencia de que podría haberlo hecho aquella vez, aunque no lo hiciera». ¿Qué clase de experimentos ulteriores podrían confirmar la creencia de Austin de que podría haberlo hecho? ¿Experimentos en el *greert*? ¿Podría verse apuntalada su creencia si embocara diez réplicas casi idénticas de aquel golpe corto? Si ésta es la clase de experimento en que está pensando, entonces no está tan interesado como pretende en las condiciones tal como se dieron. Para verlo más claro, supongamos que los «experimentos ulteriores» de Austin consisten en sacar una caja de cerillas y encender diez seguidas. «Ahí tienes —dice—, podría haber metido *ese golpe en concreto.»* Nosotros tendríamos toda la razón en objetar que dichos experimentos no guardaban la menor relación con su afirmación. Embocar diez tiros cortos tampoco

guardaría relación con su afirmación, interpretada en sentido estrecho, en cuanto afirmación respecto a las condiciones «exactamente tal como se dieron». Nosotros sugerimos que Austin haría mejor en considerar que «Austin emboca el golpe» es posible si, en situaciones *muy parecidas* a la situación en cuestión, Austin emboca el golpe. Pensamos que esto es lo que quería decir en realidad, y que haría bien en conceptualizar de este modo su golpe. Esta es la manera familiar, razonable y útil de conducir los «experimentos ulteriores» siempre que estemos interesados en comprender las causas que intervienen en un fenómeno que nos interesa. Variamos levemente (y a menudo metódicamente) las condiciones iniciales para ver qué cambia y qué permanece inalterado. Esta es la forma de reunir información *útil* del mundo para guiar nuestras campañas ulteriores de evitación y progreso.

Curiosamente, esa idea misma es la que propuso, al menos indirectamente, G. E. Moore en la obra que Austin estaba criticando en el pasaje citado. Los ejemplos de Moore eran muy sencillos: los gatos pueden trepar a los árboles y los perros no, y un barco de vapor que viaja ahora a 25 nudos puede, naturalmente, hacerlo también a 20 nudos (pero no, por supuesto, en las circunstancias *precisas* en que se encuentra ahora, con el motor en «avante a toda máquina»). El sentido del término «poder» invocado en estas afirmaciones nada problemáticas, el sentido que Honoré (1964) llamó «poder (general)» en un artículo importante pero ignorado, es tal que *nos exige* que prestemos atención no a las «condiciones exactamente tal como se dieron», sino a variaciones menores respecto a aquellas condiciones.

Así pues, las consideraciones de Austin sobre posibilidades resultan equívocas. En realidad, el método estrecho para escoger X no tiene la importancia que él y muchos otros imaginan. De ello se sigue que la verdad o la falsedad del determinismo no debe afectar a nuestra creencia de que ciertos eventos no realizados eran sin embargo «posibles», en un *sentido importante y cotidiano* de la palabra. Podemos reforzar esta idea si realizamos una visita a un dominio limitado en el que sabemos con certeza que reina el determinismo: el reino de los programas informáticos de ajedrez.

UNA MARATÓN DE AJEDREZ PARA ORDENADORES

Los ordenadores son un excelente ejemplo de los ideales deterministas de Laplace y Demócrito. Es de sobra conocido que se puede hacer que un ordenador ejecute algunos billones de pasos, para luego volverlo a si-

tuar al estado (digital) *exacto* en el que estaba antes, y ver cómo ejecuta *exactamente* los mismos billones de pasos otra vez, y otra, y otra. El mundo subatómico en el que viven los ordenadores podrá ser o no determinista, y por lo tanto también las partes subatómicas de las que están hechos éstos, pero los ordenadores en sí han sido brillantemente diseñados para ser deterministas frente al ruido microscópico e incluso al azar cuántico, ya que el hecho de ser digitales en lugar de analógicos les permite absorber estas fluctuaciones. La idea fundamental que hay detrás de digitalizar para producir determinismo es que el propio diseño nos permite *crear* hechos históricos inertes. Al clasificar forzosamente todos los eventos relevantes en dos categorías —alto o bajo; ENCENDIDO O APAGADO; 0 o 1— se garantiza que las diferencias mínimas (entre diferentes voltajes, diferentes matices de ENCENDIDO, diferentes sombras de 0) sean brutalmente descartadas. No se permite que nada oscile en ellos, y los hechos relativos a variaciones históricas que *no suponen diferencia alguna* para la serie subsiguiente de estados por los que pasa el ordenador se desvanecen sin dejar rastro.

CONRAD: ¿Los ordenadores son deterministas? ¿Se puede hacer que repitan exactamente el mismo billón de pasos una y otra vez? ¡Y qué más! Si es así, ¿por qué se cuelga tantas veces mi portátil? ¿Por qué mi procesador de texto se bloquea el martes cuando estaba haciendo la misma cosa que funcionó perfectamente el lunes?

Lo que pasa es que no estaba haciendo la *misma* cosa. Si se colgó no es porque sea indeterminista, sino porque el martes no estaba *exactamente* en el mismo estado que el lunes. Su portátil tiene que haber hecho algo en el intervalo que puso alguna «marca» oculta o activó alguna parte del procesador de textos que nunca antes había activado, y que ha cambiado un bit en alguna parte que quedó guardado en la nueva posición al apagarse el ordenador, y ahora el programa ha tropezado con aquel pequeño cambio y se ha colgado. Y si de algún modo consigue volver a ponerlo una segunda vez *exactamente* en el mismo estado que el martes por la mañana, volverá a colgarse.

CONRAD: ¿Y qué pasa con el «generador de números aleatorios»? Yo creía que mi ordenador llevaba incorporado un dispositivo para generar azar siempre que fuera preciso.

Todos los ordenadores actuales vienen equipados con un generador de números aleatorios interno que puede ser consultado siempre que sea necesario por cualquier programa activo. La secuencia de números que genera no es realmente aleatoria, sino sólo pseudoaleatoria: es «matemáticamente comprimible» en el sentido de que se trata de una secuencia infinitamente larga de números que puede ser recogida en un mecanismo de especificaciones finitas que se encarga de irlos mostrando. Cada vez que usted activa el generador de números aleatorios —cada vez que reinicia el ordenador, por ejemplo— éste ofrece exactamente la misma secuencia de dígitos, aunque se trata de una secuencia que *en apariencia* no sigue ningún patrón, como si hubiera sido generada por fluctuaciones genuinamente aleatorias. (Es más bien como una larga cinta de vídeo en la que hay grabada la historia de miles de jugadas a la ruleta. La cinta regresa siempre al «principio» cuando se reinicia el ordenador.) A veces esto puede ser un problema; los programas informáticos que recurren al azar en varios puntos de «elección» tenderán a ofrecer exactamente la misma secuencia de estados si son ejecutados una y otra vez desde el arranque del ordenador, y si queremos comprobar si un programa tiene fallos acabaremos probando siempre la misma «muestra azarosa» de estados, a menos que tomemos medidas (bastante sencillas) para obligar al programa a buscar en otro lado, de vez en cuando, dentro de su flujo de dígitos para encontrar su nuevo número «aleatorio».

Supongamos que instalamos dos programas de ajedrez diferentes en nuestro ordenador y que los conectamos a través de un pequeño programa supervisor que los obliga a jugar el uno contra el otro, partida tras partida, en una serie potencialmente infinita. ¿Jugarán la misma partida, una y otra vez, hasta que apaguemos el ordenador? *Podríamos* hacer que fuera así, pero entonces no aprenderíamos nada interesante sobre los dos programas, A y B. Supongamos que A gana a B en esta repetitiva partida. No podríamos deducir de esto que A es en general un programa mejor que B, o que A ganaría a B en una partida distinta, y la repetición exacta de la misma partida no nos permitiría aprender nada sobre las fuerzas y las debilidades respectivas de los dos programas. Sería mucho más informativo organizar un torneo para que A y B jugaran una sucesión de partidas distintas. No es muy difícil de conseguir. Si cualquiera de los dos programas consulta su generador de números aleatorios en sus cálculos (si, por ejemplo, «lanza una moneda» periódicamente para escapar a los casos en los que no encuentra ninguna razón para hacer una cosa en lugar de otra en el curso de su búsqueda heurística), en la partida siguiente el estado del

generador de números aleatorios habrá cambiado (a menos que lo arreglemos para que se reinicialice), y por lo tanto se explorarán diferentes alternativas, en un orden diferente, lo que llevará ocasionalmente a la «elección» de movimientos distintos. Florecerá una variante de la partida, y luego una tercera que será diferente en otros aspectos, lo que dará como resultado una serie en la que no habrá dos partidas iguales, como sucede con los copos de nieve. Sin embargo, si apagáramos el ordenador y luego lo reiniciáramos con el mismo programa, obtendríamos exactamente el mismo variado catálogo de partidas.

Supongamos ahora que creamos un universo de ajedrez de este tipo con dos programas distintos, A y B, y que estudiamos los resultados de una serie de mil partidas, por ejemplo. Encontraremos gran cantidad de pautas altamente fiables. Supongamos que descubrimos que, en mil partidas *distintas,* A siempre gana a B. Tal vez queramos encontrar una explicación para esa pauta y decir simplemente: «Como el programa es determinista, A estaba *predeterminado* a ganar siempre a B» no contribuirá en nada a aplacar nuestra más que razonable curiosidad. Queremos saber qué hay en la estructura, los métodos y las disposiciones de A que explica su superioridad en el ajedrez. A posee una competencia o una capacidad que B no tiene, y queremos aislar el factor interesante. Para examinar la cuestión, debemos adoptar una perspectiva de alto nivel desde la que se hagan aparentes los objetos «macroscópicos» de la toma de decisiones ajedrecísticas: representaciones de piezas de ajedrez, posiciones sobre el tablero, evaluaciones de continuaciones posibles, decisiones acerca de qué continuaciones adoptar, y así sucesivamente. O también podría ser que la explicación se encontrara a un nivel inferior; podría resultar, por ejemplo, que el programa A y el programa B fueran *idénticos* por lo que respecta a la evaluación de los movimientos de ajedrez, pero que el programa A tuviera una codificación más eficiente, de manera que pudiera llegar más lejos en sus exploraciones que el programa B en el mismo número de ciclos de la máquina. A «piensa los mismos pensamientos» que B sobre el ajedrez, pero simplemente los piensa más rápido.

En realidad, sería más interesante si no ganara siempre el mismo programa. Supongamos que A gana *casi* siempre a B y supongamos que A evalúa los movimientos empleando un conjunto distinto de principios. En tal caso tendríamos algo más interesante por explicar. Para investigar *esta* cuestión causal, deberíamos estudiar la historia de las mil partidas diferentes en busca de pautas ulteriores. Podríamos estar seguros de encontrar muchas. Algunas de ellas serán endémicas en cualquier partida de

ajedrez que se juegue (por ejemplo, la certeza casi absoluta de la derrota de B en cualquier partida en la que B tenga una torre menos) y algunas de ellas serán peculiares de A y B como jugadores particulares de ajedrez (por ejemplo, la tendencia de B a perder su reina pronto). Descubriríamos sus modelos básicos de estrategia ajedrecística, como el hecho de que cuando a B se le acaba el tiempo, examina menos a fondo las ramas restantes del árbol de posibilidades de la partida que cuando le queda más tiempo, estando en la misma posición. En resumen, encontraríamos gran abundancia de regularidades *explicativas,* algunas de las cuales no conocerían excepciones (en nuestra serie de mil partidas) mientras que otras tendrían un carácter estadístico.

Dichas pautas macroscópicas son momentos destacados en el desarrollo de un espectáculo determinista que, visto desde la perspectiva de la microcausalidad, viene a ser siempre el mismo. Lo que desde nuestra perspectiva parece el combate lleno de suspense entre dos programas de ajedrez puede verse bajo el «microscopio» (al contemplar el flujo de datos e instrucciones por la CPU del ordenador) como un sencillo autómata determinista que actúa de la única manera que puede hacerlo, con cambios siempre predecibles si se examina el estado preciso del generador de números pseudoaleatorios. No hay «auténticas» encrucijadas o ramificaciones en su futuro; todas las «elecciones» tomadas por A y B están predeterminadas. Según parece, no es *posible* que en este mundo ocurra nada distinto de lo que ocurre. Supongamos, por ejemplo, que en el tiempo *t* B se expone a un posible jaque mate, pero éste no llega a producirse porque a A se le acaba el tiempo y da por concluida su búsqueda del movimiento clave un ciclo antes de encontrarlo. Aquel jaque mate *nunca va a producirse.* (Esto es algo que podríamos demostrar, por si teníamos alguna duda, con sólo ejecutar exactamente el mismo torneo otro día. En el mismo momento de la serie, a A se le acabaría el tiempo otra vez y daría por concluida su búsqueda exactamente en el mismo punto.)

Así pues, ¿qué debemos decir? ¿Es este mundo de juguete un mundo donde no hay prevención, ni defensa, ni ataque, ni oportunidades perdidas, un mundo sin posibilidades genuínas, que no conoce el toma y daca de la agencia genuina? Es necesario admitir que nuestros programas de ajedrez, igual que los insectos o los peces, son agentes demasiado simples como para ser candidatos plausibles a un libre albedrío moralmente significativo, pero el determinismo de su mundo no les quita sus diferentes talentos, sus diferentes capacidades de aprovechar las oportunidades que se les presentan. Si queremos comprender lo que ocurre en ese mundo,

podemos —o en realidad debemos— hablar sobre los cambios que introducen en sus circunstancias a través de sus elecciones informadas, y sobre lo que *pueden* y *no pueden* hacer. Si queremos desvelar las *regularidades causales* que explican las pautas que descubrimos en aquellas mil partidas, debemos tomarnos en serio la perspectiva que describe este mundo como si contuviera dos agentes, A y B, cada uno de los cuales trata de derrotar al otro al ajedrez.

Supongamos que amañamos el programa que supervisa el torneo para que siempre que A gane suene una campana y cada vez que gane B suene un timbre. Ponemos en marcha la maratón, y un observador que no sabe nada del programa observa que la campana suena con bastante frecuencia, mientras que el timbre apenas lo hace nunca. Quiere saber qué es lo que explica esta regularidad. La regularidad con la que A vence a B puede ser advertida y descrita con independencia de si adoptamos la perspectiva intencional, y necesita una explicación. La única explicación —la explicación correcta— sería que A genera mejores «creencias» sobre lo que B hará si... de las que B genera sobre lo que A hará si... En un caso como éste, adoptar la perspectiva intencional es un *requisito* para encontrar la explicación.

Supongamos que encontramos dos partidas dentro de la serie en las que los primeros doce movimientos son iguales, pero en la primera A juega con blancas y en la segunda con negras. En el movimiento 13 de la primera partida, B «mete la pata» y, a partir de ahí, todo va cuesta abajo para A. En el movimiento 13 de la segunda partida, A, en cambio, encuentra el movimiento salvador, el enroque, y termina ganando. «B *podría haberse enrocado* en aquel punto de la primera partida», dice un observador, haciéndose eco de Austin. ¿Verdadero o falso? El movimiento, el enroque, era igual de legal la primera vez, de modo que en este sentido entraba dentro de las «opciones» disponibles para B. Supongamos que descubrimos, además, que el enroque no era sólo uno de los movimientos posibles que se había representado B, sino que B llegó a realizar una somera exploración de las consecuencias del enroque, que abandonó, por desgracia, antes de que se le revelaran sus virtudes. En tal caso, ¿*podía* haberse enrocado B? ¿Qué es lo que estamos tratando de descubrir? Mirar *exactamente* el mismo caso, una y otra vez, no resulta nada informativo: lo que lleva en realidad al diagnóstico es prestar atención a otros casos parecidos. Si descubrimos que, en otras circunstancias parecidas, B *sí es* capaz de llevar más lejos su evaluación y adivina las virtudes de movimientos como ése y los realiza efectivamente —si descubrimos, en el caso extremo, que cam-

biar un solo bit en el generador de números aleatorios daría como resultado el enroque de B—, habríamos confirmado (sobre la base de «experimentos ulteriores») la convicción del observador de que B podía haberse enrocado entonces. Diríamos que el hecho de que B no se enrocara fue pura chiripa, mala suerte con el generador de números aleatorios. Si, en cambio, comprobamos que descubrir las razones en favor del enroque requiere un análisis mucho más largo del que B puede ejecutar en el tiempo disponible (mientras que la superior potencia de A le permite realizar la tarea), tendremos base suficiente para concluir que B, a diferencia de A, no podía haberse enrocado. Tal vez descubriéramos que enrocarse era uno de esos movimientos que va seguido de un «(!)» en los problemas de ajedrez de los periódicos, un movimiento «profundo» que estaba fuera del alcance de B. Imaginar que B se enrocara requeriría demasiadas alteraciones en la realidad; estaríamos cometiendo el error antes mencionado de permitir que X abarcara demasiado.

En resumen, usar el método estrecho para escoger X no nos sirve de nada si lo que queremos es explicar las pautas que se manifiestan en los datos. Sólo obtendremos algún conocimiento si «removemos los hechos» (tal como ha dicho David Lewis) y examinamos *no* las «condiciones exactamente tal como se dieron», sino las de mundos vecinos. En cuanto ampliamos un poco X, descubrimos que B tiene opciones adicionales, en un sentido tanto informativo como moralmente relevante (si pasamos a otros mundos que van más allá del ajedrez). Muchos filósofos han dado por supuesto, sin aportar argumentos específicos para ello, que cuando preguntamos por las posibilidades que había en cierto momento, estamos —y deberíamos estar— interesados en saber si, en caso de darse *exactamente* las mismas circunstancias, volvería a producirse el mismo hecho. Hemos sostenido que, a pesar de la tendencia tradicional de los filósofos a suscribir esta estrategia, no ha sido *nunca* la que han seguido aquellos que han investigado seriamente las cuestiones relativas a la posibilidad, y no hay, en todo caso, motivo para adoptarla: *en ningún caso* se podría conseguir por esta vía una respuesta satisfactoria a nuestra curiosidad. La carga de prueba está ahora del lado de aquellos que piensan de otro modo; son ellos quienes deben explicar por qué las posibilidades «reales» requieren una elección estrecha de X... o por qué debería interesarnos una concepción de este tipo de la posibilidad, sea cual sea su «realidad».

Así pues, los mundos deterministas pueden dar cabida cómodamente a *posibilidades* en el sentido más amplio e interesante del término. En realidad, introducir el indeterminismo no añade nada a un universo en térmi-

nos de posibilidades valiosas, oportunidades o competencias. Sustituir el generador de números pseudoaleatorios por un dispositivo genuinamente indeterminista no ayudará en nada al programa B en nuestro torneo determinista de ajedrez donde siempre gana A. El programa A *seguirá* ganando cada vez. Un algoritmo superior como el de A apenas notará un cambio tan insignificante, tan invisible a la práctica. Por más que los generadores pseudoaleatorios no puedan producir resultados genuinamente aleatorios, se acercan tanto a ello que la diferencia es casi despreciable a todos los efectos. Hay sin embargo un contexto en el que sí introduce una diferencia práctica: la criptografía. Los superordenadores son capaces de dar con la clave de la aparente falta de pauta de los algoritmos particulares pseudoaleatorios de generación de números, lo que da una importancia especial al uso de números genuinamente aleatorios en tales contextos especializados.[4] Pero fuera de este contexto, donde debemos preocuparnos por la existencia de un oponente que puede tener acceso a nuestro particular modelo de generador de números pseudoaleatorios y usarlo para «leernos la mente», no tenemos nada que ganar con ser genuinamente indeterministas. Para expresarlo de forma más gráfica, el universo podría ser determinista los días pares del mes e indeterminista los días impares y nunca descubriríamos una diferencia en las oportunidades o las capacidades de los seres humanos; lograríamos los mismos éxitos —y los mismos lamentables fracasos— el 4 de octubre que el 3 o el 5 de octubre. (Si su horóscopo le hubiera aconsejado posponer cualquier decisión moralmente importante a un día impar, no tendría mayor motivo para seguir su consejo que si le dijera que esperara a la luna menguante.)

EVENTOS SIN CAUSA EN UN UNIVERSO DETERMINISTA

> *La independencia causal generalizada de ocurrencias simultáneas es lo que preserva el margen de maniobra en el Universo.*
>
> ALFRED NORTH WHITEHEAD, *Adventures of Ideas*

El determinismo es una doctrina relativa a la suficiencia: si S_0 es una proposición (inconcebiblemente compleja) que especifica con perfecto detalle la descripción del estado del universo en t_0, y si Si especifica del

4. Si las necesita, puede conseguir secuencias aleatorias de números de diversas fuentes en Internet, como en http://www.random.org y http://www.fourmilab.ch/hotbits

mismo modo la descripción del estado del universo en un momento posterior tu entonces el determinismo establece que S_0 es suficiente para Si en todos los mundos físicamente posibles. Sin embargo, el determinismo no nos dice nada sobre qué condiciones previas son *necesarias* para producir Si o cualquier otra proposición, en realidad. En consecuencia, como la causalidad presupone en general la necesidad, la verdad o falsedad del determinismo tiene poca o ninguna importancia para la validez de nuestros juicios causales.

Por ejemplo: según el determinismo, la situación exacta del universo un segundo después del Big Bang (llamemos S_0 a la proposición correspondiente) es causalmente suficiente para producir el asesinato de John F. Kennedy en 1963 (proposición C). Sin embargo, no hay razón para decir que S_0 causó C. Aunque sea una razón suficiente, no tenemos razón para creer que S_0 sea necesario. Por lo que podemos saber, Kennedy podría muy bien haber sido asesinado en cualquier caso, aunque se hubieran dado condiciones distintas en el nacimiento del universo. ¿Cómo podríamos saberlo? Podemos imaginarnos cómo sería la investigación, aunque no podamos llevarla a cabo: imaginemos que tomamos una instantánea del universo en el momento del asesinato de Kennedy y luego alteramos la imagen en algún aspecto trivial (pongamos que movemos a Kennedy 1 milímetro hacia la izquierda). La proposición C: «John F. Kennedy fue asesinado en 1963 (en Dealey Plaza, cuando iba en su coche oficial...)» sigue siendo cierta, pero con una diferencia microscópica en las condiciones atómicas que la convierten en verdadera. Luego, partiendo de nuestra descripción de estado sutilmente revisada de 1963 y siguiendo las leyes (deterministas) de la física a la inversa, generamos una película que recorre toda la historia hasta el Big Bang, con lo que obtenemos un mundo en el que S_0 no se aplica por una sutil diferencia. Hay mundos posibles muy parecidos al nuestro donde Kennedy muere pero S_0 no se cumple, de modo que el estado del universo descrito por S_0 *no* es la causa del asesinato de Kennedy. Otras más plausibles de aquel evento incluirían: «Una bala sigue un curso que se dirige al cuerpo de Kennedy»; «Lee Harvey Oswald apretó el gatillo de su arma». Los grandes ausentes de esta lista son las descripciones microscópicas del universo billones de años antes del incidente. Los filósofos que afirman que según el determinismo S_0 «causa» o «explica» C pasan por alto la idea central de la investigación causal, y éste es el segundo error capital.

En realidad, el determinismo es perfectamente compatible con la noción de que algunos eventos no tienen ninguna causa. Consideremos la proposi-

ción: «La devaluación de la rupia hizo que se desplomara el índice Dow Jones». Con razón miramos con recelo esta declaración; ¿estamos realmente seguros de que dentro del conjunto de los universos vecinos al nuestro el Dow Jones se desplomó *únicamente* en aquellos donde la rupia había caído primero? ¿Podemos imaginarnos siquiera que en todos los universos donde cayera la rupia se dispararan las órdenes de venta en el mercado de valores? ¿No es posible que se diera una confluencia de multitud de factores que bastaran en conjunto para hacer que se desplomara el mercado, sin que ninguno de ellos fuera esencial en sí mismo? Tal vez haya días en los que se pueda encontrar una razón válida para el comportamiento de Wall Street; pero al menos con igual frecuencia sospechamos que no hay ninguna causa particular que lo explique.

El lanzamiento de una moneda que no esté trucada es un ejemplo familiar de un evento que tiene un resultado (por ejemplo, cara) que propiamente *no tiene causa*. No tiene causa porque no importa cómo escojamos el conjunto X (ignorando el consejo erróneo de Austin de considerar las circunstancias tal como se dieron *exactamente*), no encontraremos ningún rasgo de C que sea necesario para que salga cara o para que salga cruz. ¿Ha pensado usted alguna vez en la aparente contradicción que supone usar una moneda como forma de generar un evento aleatorio? No hay duda de que el resultado del lanzamiento de una moneda al aire es la consecuencia *determinista* de la suma total de las fuerzas que actúan sobre ella: la velocidad y la dirección del lanzamiento que imparte el giro, la densidad y la humedad del aire, la rotación de la Tierra, la distancia entre Marte y Venus en aquel momento, y así sucesivamente. La cuestión es que esta suma total no contiene pautas predictivas. Tal es el truco que hay **de** trás de los mecanismos generadores de aleatoriedad del tipo de lanzar una moneda al aire: convertir en incontrolable el resultado al hacerlo sensible a tantas variables que no podemos señalar como causa ninguna lista finita o factible de condiciones. Ése es el motivo por el que pedimos que se lance la moneda bien alta, con mucho efecto, y no simplemente que se deje caer de los dedos una pulgada por encima de la mesa: ponemos en marcha una secuencia que prácticamente garantiza que nada será la causa de que el resultado sea cara o cruz. Nótese que la estrategia de lanzar la moneda utiliza la digitalización para garantizar que el resultado carece de causa (si se hace debidamente). Cumple exactamente la función contraria de la que cumple la digitalización en los ordenadores: en lugar de absorber todas las microvariaciones del universo, las amplifica, y garantiza que la suma inimaginablemente larga de fuerzas que actúan en aquel momen-

to situará el digitalizador en uno de sus dos estados, cara o cruz, pero sin que haya condiciones necesarias destacables para ninguno de los dos resultados.

La práctica de «remover los hechos» en experimentos controlados es una de las grandes innovaciones de la ciencia moderna y, tal como señala Judea Pearl, depende de que usemos algo como el lanzamiento de una moneda para *romper* los lazos causales que de otro modo existen entre los eventos que queremos analizar:

> **Supongamos que queremos estudiar el efecto de cierto tratamiento farmacológico para la recuperación de pacientes aquejados de cierta enfermedad [...]. Bajo condiciones no controladas, la elección del tratamiento está en manos de los pacientes y puede depender de su trasfondo socioeconómico. Esto crea un problema, ya que no podemos decir si los cambios en los índices de recuperación son debidos al tratamiento o a dichos factores de fondo. Lo que nos interesa es comparar pacientes de trasfondos parecidos, y eso es precisamente lo que consigue el *experimento aleatorio* de [sir Ronald] Fisher. ¿Cómo? En realidad consta de dos partes: aleatorización e intervención.**
>
> **La intervención significa que alteramos el comportamiento natural del individuo: separamos a los sujetos en dos grupos, llamados tratamiento y control, y convencemos a los sujetos para que sigan las reglas del experimento. Asignamos tratamiento a algunos pacientes que, bajo circunstancias normales, no buscarían tratamiento, y damos placebo a los pacientes que sí recibirían tratamiento en otro caso. Eso, en nuestro nuevo vocabulario, es hacer cirugía: cortamos un vínculo funcional y lo reemplazamos por otro. La gran idea de Fisher fue que conectar el nuevo vínculo con el lanzamiento aleatorio de una moneda *garantiza* que el vínculo que deseamos romper se rompe realmente. La razón es que se presume que una moneda lanzada al azar no se ve afectada por nada que podamos medir a nivel macroscópico (incluido, por supuesto, el trasfondo socioeconómico de un paciente) (Pearl, 2000, pág. 348).**

Nuestra práctica en estos casos revela una presunción de fondo que parece estar ampliamente aceptada (aunque nunca o raras veces es sometida a examen): la presunción de que la única forma de que un evento no tenga causa es que se encuentre estrictamente subdeterminado, que *no* tenga condición suficiente, por más difusa, compleja y poco interesante que sea ésta. Esto puede llevar a graves distorsiones de nuestra agenda científica: ¿cuál fue la causa de la Primera Guerra Mundial? ¡Si queremos ser unos buenos investigadores científicos, lo menos que podemos hacer es encontrarle una causa! Declarar que la Primera Guerra Mundial no tuvo ninguna causa vendría a ser lo mismo que declararla o bien una violación

de las leyes de la naturaleza —¡un milagro!— o bien (la física cuántica al rescate) el resultado de procesos cuánticos indeterministas. Pues bien, no es así. *Podría* ser que por más que los historiadores «removieran los hechos» en busca de los antecedentes necesarios de la Primera Guerra Mundial en mundos posibles cercanos, se encontraran con que aquellos universos donde se produce la Primera Guerra Mundial no comparten ningún antecedente común y necesario. Supongamos, por ejemplo, que en el universo A, el archiduque Fernando es asesinado y se declara subsiguientemente la Primera Guerra Mundial. ¿Significa eso que lo anterior es la causa de esta última (tal como algunos de nosotros «aprendimos» en la escuela)? Tal vez no; tal vez en un universo B, el archiduque Fernando sobrevive, pero la Primera Guerra Mundial estalla igualmente. Y, del mismo modo, podría darse el caso de que para cada «causa» que propusiera el historiador X, el historiador Y fuera capaz de imaginar un mundo en el que se produce la Primera Guerra Mundial sin venir precedida por dicha candidata a causa. La guerra podría haber sido un golpe de azar, en cuyo caso persistir en disputas sobre cuál fue «la causa» no sólo sería inútil, sino que casi garantizaría la creación de mitos artificiales sobre causas secretas que explorar. La búsqueda de tales condiciones necesarias es siempre racional, mientras tengamos presente que *puede* ser que en algún caso particular no haya nada que encontrar.[5]

Alguien podría preguntarse por qué nos preocupa tanto la necesidad causal. Volvamos por un momento a los programas de ajedrez A y B. Supongamos que nos llama la atención una de las raras partidas en las que B gana, y que queremos saber «la causa» de esta sorprendente victoria. La tesis trivial de que la victoria de B la «causó» el estado inicial del ordenador no resultaría nada informativa. Por supuesto, el estado total del universo del modelo en los momentos anteriores era *suficiente* para que se produjera la victoria; pero lo que queremos saber es qué rasgos eran *necesarios*, para así comprender qué tienen en común estos raros eventos. Queremos descubrir cuáles son estos rasgos, la ausencia de los cuales traería de manera más directa la derrota de B, el resultado normal. Tal vez

5. La tendencia no sólo a buscar, sino también a encontrar causas no es ociosa, tal como observa Matt Ridley en su estudio de la enfermedad Creutzfeldt-Jakob, para la que no se ha encontrado aún causa alguna: «Esto ofende nuestro determinismo natural, de acuerdo con el cual las enfermedades deben tener causas. Pero cabe la posibilidad de que la ECJ simplemente surja de manera espontánea en una proporción de un caso por millón al año» (Matt Ridley, 1999, pág. 285).

descubramos un defecto hasta ahora inadvertido en la estructura de control de A, un fallo que no se ha manifestado hasta ahora. O tal vez descubramos un rasgo idiosincrásico de brillantez en la competencia de B, que una vez diagnosticado nos permitiría predecir exactamente qué circunstancias harían posible en el futuro otra victoria parecida de B. O tal vez la victoria es una coincidencia de múltiples condiciones que no justifican ninguna reparación, pues la probabilidad de que se reproduzcan es nula en la práctica. Esta última posibilidad, es decir, que en un sentido relevante simplemente no hubiera causa para la victoria de B —que fuera un golpe de suerte—, resulta fácil de comprender en un contexto simplificado como éste, pero es más difícil de admitir, según parece, en casos reales.

La racionalidad *requiere* que evaluemos las condiciones necesarias al menos con el mismo cuidado que las condiciones suficientes. Consideremos el caso de un hombre que cae por el hueco del ascensor. Aunque no sabe exactamente en qué mundo posible habita, sí sabe una cosa: se encuentra dentro de un conjunto de mundos en *todos* los cuales en breve acabará en el fondo del hueco del ascensor. La gravedad se encargará de eso. El batacazo es en este sentido *inevitable,* porque se produce en todos los mundos consistentes con lo que conoce. Pero tal vez no sea inevitable el hecho de *morir.* Cabe la posibilidad de que en alguno de los mundos donde se pega el batacazo logre, sin embargo, sobrevivir. Dichos mundos no incluyen ninguno donde caiga, por decir algo, con la cabeza por delante o con los brazos abiertos como un pájaro, pero puede haber mundos en los que caiga con los pies por delante, amortigüe el golpe y sobreviva. Hay cierto margen. Puede planificar racionalmente sus acciones partiendo de la premisa de que es posible que viva e, incluso aunque no pueda *descubrir* condiciones suficientes que garanticen su supervivencia, puede al menos mejorar sus opciones al realizar todas las acciones que sean necesarias y, por lo tanto, con un poco de suerte, encontrarse en el Vasto número de mundos posibles en los que vive.

> CONRAD: Una vez más, ¿qué sentido tiene hablar de mejorar sus opciones? Estamos presuponiendo el determinismo. No puede *cambiar de mundo.* Está en el mundo en el que está, el mundo presente, y en este mundo vive o muere, y ahí termina la cuestión.

De acuerdo, pero eso es cierto con independencia del determinismo y es irrelevante para la cuestión de la racionalidad de su acción. Supongamos que suspendemos temporalmente su caída y *que* le permitimos leer el

Vasto rincón de la Biblioteca de Babel donde se encuentran las biografías de un hombre que responde a su mismo nombre, posee sus mismos rasgos, características e historia hasta la fecha: la historia de un hombre que cae accidentalmente por el hueco de un ascensor y se encuentra frente a una colección inimaginablemente grande de libros, cada uno de los cuales pretende ser la verdadera historia de su vida. En algunos de estos libros vive y en otros muere (y, tratándose de la Biblioteca de Babel, en algunos de ellos se convierte en una taza de oro y es arrojado a Cleopatra por un caracol gigante). El problema es que aunque puede descartar los libros fantásticos sobre la base de su conocimiento general de cómo funciona el mundo, no tiene ninguna forma de saber qué libro en concreto es el verdadero entre los que dicen que vive o muere tras su caída. Y suponer que el determinismo sea verdadero, *o falso,* no le ayudará a encontrar la aguja en este pajar. La mejor estrategia que puede adoptar frente a esta incertidumbre insuperable respecto a cuál de los libros dice la verdad es buscar pautas generales predictivas —causas y efectos— y dejarse guiar por las predicciones que sugieren. Pero ¿cómo va a hacer eso? No hay problema: eones de evolución le han diseñado para hacer precisamente eso. Si no tuviera estos talentos, no estaría allí. Es el producto de un proceso de diseño que ha creado una especie de previsores-evitadores en quienes este truco es como una segunda naturaleza. No son perfectos, pero obtienen resultados mucho mejores que el mero azar. Comparemos, por ejemplo, las perspectivas de unos seres que tienen ante sí la oportunidad de ganar un millón de dólares con un lanzamiento de moneda o una jugada a los dados en la que salgan dos unos. Algunos de ellos razonan en términos fatalistas: «No importa qué método escoja; las opciones de que salgan dos unos son 0 o 1. No sé cuál es el destino ya predeterminado, y lo mismo sucede con la moneda». Otros actúan guiados por la convicción de que la probabilidad de 1 entre 2 que tienen al escoger la moneda es mucho mejor que la probabilidad de 1 entre 36 de que les salga el doble uno, y optan por lanzar la moneda. No es ninguna sorpresa que la gente diseñada para hacerlo así haya obtenido mucho mejores resultados que los fatalistas, los cuales puede decirse que tienen un defecto de diseño, desde la perspectiva que da la historia.

¿SERÁ EL FUTURO IGUAL QUE EL PASADO?

Y ahora, por fin, estamos listos para enfrentarnos al tercer gran error en relación con el determinismo. Algunos pensadores han planteado que

la verdad del determinismo implicaría uno o más de los siguientes desalentadores corolarios: todas las tendencias son permanentes, el carácter es en gran medida inmutable, y es improbable que uno pueda cambiar en el futuro la propia manera de ser, la propia fortuna, o la propia naturaleza básica. Ted Honderich, por ejemplo, ha sostenido que el determinismo destruye de algún modo lo que llama nuestras esperanzas vitales:

> **Si a una persona le han ido bien las cosas, todavía puede esperar más del futuro partiendo de la presunción de que su vida entera está prefijada [...]. Si no le han ido bien las cosas, o no tan bien como esperaba, puede decirse cuando menos que no es insensato que tenga mayores esperanzas partiendo de la presunción de que el conjunto de la propia vida no está prefijado, sino que está vinculado a la actividad del yo [...]. Dada la premisa optimista de nuestra razonabilidad, hay razones para pensar que *no* tendemos a la idea de un futuro personal prefijado (Honderich, 1988, págs. 388-389).**

No hay duda de que tales inquietudes tienen su origen en la vaga noción de que las verdaderas posibilidades (de mejorar nuestra fortuna, por ejemplo) desaparecen con el determinismo. Pero esto es un error. La distinción entre ser un ente con un futuro abierto y ser un ente con un futuro cerrado es estrictamente independiente del determinismo. En general, no hay paradoja alguna en la observación de que ciertos fenómenos están predeterminados para ser alterables, caóticos e impredecibles, un hecho evidente e importante que los filósofos curiosamente han ignorado. A Honderich le parece inquietante la idea de que podamos tener un <<*futuro* personal prefijado», pero las implicaciones de esta idea son enteramente distintas de las implicaciones de tener una «*naturaleza* personal prefijada». El propio futuro personal «prefijado» —es decir, predeterminado— podría consistir perfectamente en la bendición de una *naturaleza* proteica, altamente sensible a la «actividad del yo». El conjunto total de los futuros personales, esté o no «prefijado», contiene toda clase de escenarios agradables, entre los cuales habrá victorias sobre la adversidad, superaciones de la debilidad, reformas del carácter e incluso cambios de fortuna. Tan determinado podría ser el hecho de que los perros viejos *pudieran* aprender trucos nuevos como que no pudieran. La cuestión que debemos preguntar es: ¿son los perros viejos el tipo de ente al que se le pueden enseñar trucos nuevos? Si no lo son, no queremos ser como los perros viejos. Es justo que nos preocupemos por si somos la clase de entes cuya trayectoria futura no repetirá las mismas pautas del pasado, y la tesis general del determinismo no tiene ninguna implicación respecto a estas cuestiones.

Consideremos los simples mundos deterministas Vida. En un nivel nada cambia nunca; los píxels hacen lo mismo una y otra vez eternamente, siguiendo la sencilla ley de la física. En otro nivel, vemos diferentes clases de mundos. Algunos mundos son tan inmutables a vista de pájaro como lo son al nivel atómico, un campo de naturalezas muertas y semáforos, pongamos por caso, que parpadean para toda la eternidad. No hay drama, ni suspense. Otros mundos «evolucionan» constantemente, sin pasar dos veces por el mismo estado, sea de una forma ordenada, con un crecimiento predecible, al crear por ejemplo un flujo constante de planeadores idénticos y separados por el mismo espacio, o bien de una forma aparentemente desordenada, con miríadas de masas de píxels que crecen, se mueven y colisionan entre sí. En esos mundos, ¿es el futuro igual que el pasado? Sí y no. La física es eternamente inmutable, de modo que los microeventos son siempre iguales. Pero, a un nivel superior, el futuro puede ser muy variado: puede contener pautas que sean iguales que las del pasado y otras que sean completamente nuevas. Es decir, que en algunos mundos deterministas hay cosas cuya *naturaleza* cambia con el tiempo, de modo que el determinismo no implica una naturaleza fija. Un hecho pequeño, pero alentador. Y aún nos faltan uno cuantos más.

Algunos mundos Vida contienen competiciones y, aunque el demonio de Laplace sabe exactamente cómo terminará cada competición, puede resultar auténticamente dramático e inquietante para inteligencias inferiores, que no pueden saber, desde su limitada perspectiva, cómo terminará la competición. Consideremos, por ejemplo, aquellos mundos Vida en los que hay una Máquina Universal de Turing que ejecuta nuestro programa de partidas de ajedrez de A contra B. El ajedrez es un juego de «información perfecta»; en este sentido es distinto de los juegos de cartas, en los que se ocultan las cartas al oponente (y en los que ningún oponente sabe qué carta saldrá a continuación de la baraja). Así pues, A y B tienen una información completa y compartida sobre el estado de la partida de ajedrez en curso y sobre las posibilidades que se abren a continuación. Sin embargo, acaban creando como resultado de sus esfuerzos diferentes inventarios de expectativas acerca de los probables movimientos futuros de sus oponentes, y de ellos mismos. La competición consiste en utilizar la información compartida para generar una información privada sobre la que basar la elección de los propios movimientos, y la *explicación* de por qué A gana a B (si lo hace, y cuando lo hace) debe formularse en términos de su superior capacidad para generar, y usar, información acerca de un futuro incierto y abierto (desde su perspectiva). Cada usuario finito de infor-

mación tiene un horizonte epistémico; no lo sabe todo del mundo que habita, y esta ignorancia insuperable garantiza que tenga un futuro *subjetivamente* abierto. El suspense es una condición necesaria de la vida para cualquier agente de este tipo.[6]

Pero dejemos a un lado la cuestión del suspense subjetivo, y del cambio de naturaleza. ¿Qué pasa con la *superación*? ¿Puede haber no sólo superación, sino una superación causada por el propio agente en un mundo determinista? ¿Puede un agente de un mundo determinista tener una esperanza realista de prosperar? Una vez más, la respuesta a esta pregunta no tiene nada que ver con el determinismo y mucho que ver en cambio con el *diseño*. Los programadores han demostrado ya que algoritmos informáticos deterministas pueden adaptarse a cambios del entorno y aprender de sus errores. Hemos dejado para más adelante la posibilidad de atribuir capacidad de aprendizaje a los programas A y B, por no querer distraer la atención de otras cuestiones que estaban en discusión, pero consideremos lo que ocurre cuando incorporamos la capacidad de aprender de la experiencia en uno de los oponentes. Si el inicialmente mediocre B poseyera la capacidad de aprender y A no, podría ser que al final B saliera victorioso. Uno de los productos de la historia de las competiciones de B, podríamos decir que el fruto de sus esfuerzos, podría ser el desarrollo de una estructura que le diera una competencia mejor y, por lo tanto, una mejor posición en la vida. B pasaría de ser el eterno perdedor al ganador habitual. Supongamos que B posee esta clase de estructura capaz de aprender en un mundo determinista; su envidiable capacidad no experimentará ninguna mejora con la introducción de un generador de números aleatorios genuínamente indeterminista. La introducción del indeterminismo tampoco contribuirá a hacer más abierto el futuro de B, si le falta esta capacidad de aprender.

Las condiciones bajo las que se produce la superación personal (a falta de milagros) son precisamente las condiciones bajo las cuales algo —sea un dios-*hacker,* o la evolución, o el instructor de B, o el propio B—

6. El demonio de Laplace ejemplifica un interesante problema planteado por primera vez por Turing, y tratado por Ryle (1949), Popper (1951) y MacKay (1960). Ningún sistema de procesamiento de la información puede disponer de una descripción completa de sí mismo: es el problema de Tristram Shandy acerca de cómo representar la representación de la representación de... los detalles más pequeños e insignificantes. En consecuencia, incluso el demonio de Laplace tiene un horizonte epistémico y, como resultado, no puede predecir sus propias acciones del mismo modo que puede predecir el siguiente estado del universo (del que debe estar necesariamente fuera).

identifica las causas que hay detrás de la victoria y genera diseños que aumentan las probabilidades de que dichas causas estén presentes en el futuro en los momentos adecuados. Hay, pues, una razón familiar para diseñar un programa que aprenda de la experiencia: en el futuro puede encontrarse con una situación parecida, y lo que ocurra entonces podría verse influido por lo que aprenda ahora. Esto es así porque lo que ocurra entonces dependerá de lo que decida entonces; si enroca o no, por ejemplo, *dependerá de él* en un sentido importante. No dependerá de él si las reglas del ajedrez siguen siendo las mismas, como tampoco los movimientos de su oponente; sus propios movimientos, en cambio, dependerán de él en el sentido que nos interesa: serán el resultado de sus *propios* procesos exploratorios y deliberativos.

En este sentido, comparemos el caso de un pez enfrentado a un anzuelo con cebo y el de un pez enfrentado a una red que se le viene encima a gran velocidad; que el primer pez muerda el anzuelo es algo que depende de sí mismo, mientras que el hecho de que el segundo pez entre en la red probablemente no. Así pues, ¿son libres los peces? No en un sentido moralmente relevante, pero sí tienen sistemas de control que toman «decisiones» de vida o muerte, lo cual es al menos una condición necesaria para la libertad. En el capítulo 4 consideraremos si hay otro sentido más profundo de «depender de» que se nos pueda aplicar a nosotros (como agentes morales) pero no a los ordenadores deterministas que juegan al ajedrez, o a los peces.

Vivimos en un mundo que está subjetivamente abierto. Y hemos sido diseñados por la evolución para ser «informóvoros», seres epistémicamente hambrientos, buscadores de información, entregados a la interminable tarea de mejorar nuestro conocimiento del mundo, para poder tomar mejores decisiones respecto a nuestro futuro subjetivamente abierto. La luna está hecha de la misma materia que nosotros, obedece a las mismas leyes de la física, pero su naturaleza, *a diferencia* de la nuestra, está fijada. Es más, a diferencia de nosotros, no se preocupa por su naturaleza. No está preparada para preocuparse lo más mínimo por sí misma. La diferencia entre nosotros y la luna no es una diferencia que tenga que ver con la física; es una diferencia de diseño de un nivel superior. Somos el producto de un proceso competitivo a gran escala para la mejora de diseños; la luna no. Este proceso de diseño, la selección natural, se basa como es sabido en la mutación «aleatoria» como Generador de Diversidad último. Hemos visto que los programas informáticos —y en general los experimentos controlados— utilizan esta clase de generadores de diversidad en

buena medida con el mismo objeto: para introducir nuevos modelos en los procesos indagatorios y hacer que superen los antiguos. Pero también hemos visto que esta bienvenida fuente de diversidad no tiene por qué ser verdaderamente aleatoria, en el sentido de ser *indeterminista.*

Decir que si el determinismo es verdadero, nuestro *futuro* está fijado, es decir... nada interesante. Decir que si el mundo es determinista, nuestra *naturaleza* está fijada, es decir algo falso. Nuestras naturalezas no están fijadas porque hemos evolucionado hasta convertirnos en entidades *diseñadas* para cambiar su naturaleza en respuesta a las interacciones con el resto del mundo. Toda la angustia acerca del determinismo tiene su origen en una confusión entre lo que supone tener una *naturaleza* fijada y tener un *futuro* fijado. La confusión surge cuando uno trata de mantener dos perspectivas simultáneas sobre el universo: la perspectiva del «ojo de Dios», ante la que se despliegan al mismo tiempo pasado y futuro, y la perspectiva situada de un agente integrado *en* el universo. Desde la perspectiva intemporal de Dios nada cambia jamás —la historia entera del universo se despliega «de una vez»— e incluso un universo indeterminista no sería sino un árbol de ramificaciones estáticas. Desde la perspectiva del agente situado, las cosas cambian con el tiempo, y los agentes cambian para hacer frente a dichos cambios. Pero por supuesto no *todos* los cambios son posibles para nosotros. Hay cosas que podemos cambiar y cosas que no podemos cambiar, y algunas de estas últimas son deplorables. Hay muchas cosas que están mal en nuestro mundo, pero el determinismo no es una de ellas, incluso aunque nuestro mundo esté determinado.

Así pues, una vez superado el miedo al determinismo físico, podemos dirigir nuestra atención al nivel biológico, desde el que tal vez podamos explicar cómo es posible que *seamos* libres, cuando otras entidades de nuestro mundo, hechas del mismo tipo de materia que nosotros, no lo son en absoluto. Como sucede habitualmente cuando se habla de biología, encontraremos que hay toda clase de categorías y grados diferentes de libertad. La libertad, tal como se manifiesta en un ordenador capaz de jugar al ajedrez que habita en el plano Vida, no es más que un juego, una mera caricatura de la clase de libertad que nos interesa. Pero estamos *realmente* interesados en esta clase de libertad, y resulta útil comenzar con el modelo más sencillo imaginable para confirmar si es compatible con el determinismo.

CONRAD: De acuerdo, ha demostrado usted que Austin estaba equivocado. Pero resulta que a él no le interesaba en absoluto cuáles eran las posibilidades *reales*; ¡lo que le interesaba era su tiro al hoyo! Y tiene razón en que la

forma de comprobar eso es hacer algunos tiros y ver cuántos entran. Tal como usted sabe, hay un sentido de competencia, de *poder hacer,* que se aplica igualmente bien a los agentes humanos y a artilugios como los ordenadores que juegan al ajedrez (y a los abrelatas, si a eso vamos). Pero todo lo que esto demuestra es que responder a *esa* clase de pregunta no roza siquiera la pregunta que a mí me interesa: ¿podría Austin haber metido *aquel tiro en concreto*? Y la respuesta a esta pregunta debe ser «no» en un mundo determinista.

Muy bien, si insiste. Tal vez haya un sentido de «posible» en el que Austin no podía haber metido aquel tiro, si el mundo fuera determinista. Ahora bien, ¿por qué razón habría de preocuparnos esta cuestión? Aparte de una ociosa curiosidad metafísica, ¿qué interés puede tener para nosotros si Austin podía o no haber metido aquel tiro en *el sentido que usted dice?*

Los incompatibilistas tienen una respuesta a esta pregunta, y antes de que podamos volver cómodamente a la evolución, deberíamos darles la oportunidad de presentarla. El siguiente capítulo está dedicado a examinar la mejor respuesta que han dado hasta el momento. Aquellos que ya estén convencidos de que el determinismo simplemente no es la cuestión pueden pasar directamente al capítulo 4, pero se perderán algunos descubrimientos incidentales sobre la naturaleza de nuestra libertad que son hasta cierto punto independientes de la defensa del indeterminismo que los reveló.

Capítulo 3

Nuestra manera cotidiana de pensar en relación con la posibilidad, la necesidad y la causalidad parece entrar en conflicto con el determinismo, pero se trata de una ilusión. El determinismo no implica que, hagamos lo que hagamos, no podíamos haber hecho otra cosa, que todo evento tiene una causa, o que nuestra naturaleza está fijada.

Capítulo 4

Un examen constructivo de un ambicioso modelo indeterminista de toma de decisiones revela tanto los motivos como los problemas que rondan a cualquier teórico que vaya por aquel camino. Se puede dar respuesta a las tesis más plausibles de los libertaristas sin necesidad de suscribir el indeterminismo, y el indeterminismo no introduce nada que pueda suponer una diferencia moral.

NOTAS SOBRE FUENTES Y LECTURAS COMPLEMENTARIAS

El libro de Judea Pearl *Causality: Models, Reasoning, and Inference* (2000), que descubrí mientras preparaba el último borrador del presente libro, plantea cuestiones acerca del modelo Taylor/Dennett en términos de mundos posibles, al tiempo que abre alternativas tentadoras. No será una tarea fácil digerirlas y, en caso necesario, reformular nuestras conclusiones, que en nuestra opinión no se ven directamente cuestionadas. Es una tarea que queda para el futuro.

Para leer más acerca de la posibilidad, véase *La peligrosa idea de Darwin* (Dennett, 1995), capítulo 5, «Lo real y lo posible» y, especialmente, «La posibilidad naturalizada» (págs. 118-123). Véase también el experimento mental («Dos Cajas Negras», págs. 412-422), que deja claro que los científicos podrían tener un *conocimiento completo* de los procesos microcausales presentes (de manera determinista) en este fenómeno y sin embargo ser totalmente incapaces de explicar las regularidades macrocausales que observan y desean explicar.

Para leer más sobre números pseudoaleatorios y su utilidad para el control y la libertad, véase *Elbow Room* (Dennett, 1984), págs. 66-67, y el resto del libro.

Publicada en nueve volúmenes entre 1759 y 1766, la novela cómica de Lawrence Sterne *Tristram Shandy* pretende ser una autobiografía, pero se enreda en espirales recurrentes de reflexión, reacción y metarreacción que la convierten en una tarea inacabada e inacabable.

NOTAS SOBRE FUENTES Y LECTURAS COMPLEMENTARIAS

El libro de Judea Pearl *Causality: Models, Reasomng, and Inference* (2000), que descubrí mientras preparaba el último borrador del presente libro, plantea cuestiones acerca del modelo Taylor/Dennett en términos de mundos posibles, al tiempo que abre alternativas tentadoras. No será una tarea fácil digerirlas y, en caso necesario, reformular nuestras conclusiones, que en nuestra opinión no se ven directamente cuestionadas. Es una tarea que queda para el futuro.

Para leer más acerca de la posibilidad, véase *La peligrosa idea de Darwin* (Dennett, 1995), capítulo 5, «Lo real y lo posible» y, especialmente, «La posibilidad naturalizada» (págs. 118-123). Véase también el experimento mental («Dos Cajas Negras», págs. 412-422), que deja claro que los científicos podrían tener un *conocimiento completo* de los procesos microcausales presentes (de manera determinista) en este fenómeno y sin embargo ser totalmente incapaces de explicar las regularidades macrocausales que observan y desean explicar.

Para leer más sobre números pseudoaleatorios y su utilidad para el control y la libertad, véase *Elbow Room* (Dennett, 1984), págs. 66-67, y el resto del libro.

Publicada en nueve volúmenes entre 1759 y 1766, la novela cómica de Lawrence Sterne *Tristram Shandy* pretende ser una autobiografía, pero se enreda en espirales recurrentes de reflexión, reacción y metarreacción que la convierten en una tarea inacabada e inacabable.

Capítulo 4

Una audiencia para el libertarismo

El problema tradicional de la libertad surge a partir de la proposición de que *si el deterninismo es verdadero, entonces no somos libres.* Esta proposición expresa el *incompatibilismo,* y ciertamente resulta plausible a primera vista. Muchas personas que han reflexionado seriamente sobre la cuestión siguen pensando que es verdadera, de modo que antes de volver a mi proyecto, que lo rechaza de plano, será bueno examinarla un poco para ver en qué consiste su atractivo, y cuáles son sus puntos fuertes, así como sus puntos débiles.

EL ATRACTIVO DEL LIBERTARISMO

Si aceptamos la proposición tal como está formulada, se nos abren dos caminos, dependiendo de qué mitad de la proposición escojamos:

> *Deterninismo duro:* el determinismo es verdadero, luego no somos libres. Los científicos de la línea más dura proclaman a veces esta postura, que declaran incluso obvia. Muchos de ellos añadirían: y si el determinismo fuera falso, *seguiríamos* sin ser libres (no somos libres en ningún caso; es un concepto incoherente). Sin embargo, a menudo evitan entrar en la cuestión de cómo justifican las firmes convicciones morales que siguen usando como guía para sus vidas. ¿Dónde nos lleva esto? ¿Qué sentido podemos encontrar en los esfuerzos, los méritos y las culpas de los seres humanos? Ya vimos en el capítulo 1 el abismo que se abre ante nosotros exactamente en este punto. ¿Hay alguna alternativa estable a esta amenaza de nihilismo moral? (Tal vez los deterministas duros que pueda haber entre los lectores descubran en capítulos ulteriores que su posición debidamente *ponderada* es que mientras que la libertad —tal como entienden ustedes el término— no existe en realidad, sí existe algo *muy parecido* a la libertad, y que es justo lo que necesitaban

para revitalizar sus convicciones morales y permitirles establecer las distinciones que necesitan establecer. Y es posible que esta clase de versión moderada del determinismo duro ya sólo sea terminológicamente distinta del *compatibilismo,* la idea de que la libertad y el determinismo son a fin de cuentas compatibles, que es la idea que defiendo en este libro.)

Libertarismo: somos realmente libres, de modo que el determinismo debe ser falso; lo que se verifica es el indeterminismo. Gracias a la física cuántica, el punto de vista más extendido entre los científicos hoy en día es que el indeterminismo es efectivamente verdadero (en el nivel subatómico y, por extensión, a niveles superiores en diversas condiciones especificables), lo que podría hacer pensar que estamos ante un final feliz para el problema; sin embargo, hay una pega: ¿cómo podemos servirnos del indeterminismo de la física cuántica para ofrecer una representación clara y coherente de un ser humano en el ejercicio de su maravillosa libertad?

Este sentido del *libertarismo*, por cierto, no tiene nada que ver con el sentido político del término. Probablemente haya más filósofos de izquierdas que de derechas en favor de esta clase de libertarismo, pero eso es sólo porque probablemente haya más filósofos de izquierdas en general. Puede ser que las personas políticamente de derechas que han pensado sobre ello tiendan a inclinarse por el libertarismo y que los conservadores religiosos se sientan atraídos hacia él, aunque sólo sea por repulsión hacia todas las alternativas, pero los libertaristas del libre albedrío no están comprometidos con ninguna idea particular sobre la relación entre los poderes del Estado y los ciudadanos. Coinciden en que la libertad requiere el indeterminismo, pero hay gran división entre ellos respecto al inconveniente antes señalado: ¿cómo puede el indeterminismo sub-atómico producir una voluntad libre? Un grupo se limita a declarar que éste no es su problema, sino el de los neurocientíficos, tal vez, o el de los físicos. Todo cuanto les preocupa es lo que podríamos llamar las restricciones de arriba a abajo que impone la responsabilidad moral: para que un agente humano pueda ser considerado propiamente responsable de algo que ha hecho, debe verificarse de algún modo que la elección de la acción por parte del agente no haya estado determinada por el conjunto de condiciones físicas que se daban antes de la elección. «Nosotros los filósofos nos encargamos de establecer las especificaciones para que un agente sea libre; dejamos el problema de la *implementación* de estas especificaciones a los neuroingenieros.» Otro grupo más reducido ha advertido que esta división del trabajo no siempre es una buena idea; la co-

herencia misma de las especificaciones libertaristas queda puesta en duda por las dificultades que surgen cuando se trata de implementarlas. Por otro lado, parece ser que el intento de desarrollar una explicación positiva de la capacidad de elección indeterminista en los seres humanos reporta beneficios que son independientes de la premisa del indeterminismo.

El mejor intento realizado hasta el momento en este sentido ha sido el de Robert Kane en su libro de 1996 *The Significance o/Free Will.*[1] Sólo una explicación libertarista, sostiene Kane, puede aportar el elemento al que todos —o al menos algunos— aspiramos y al que Kane da el nombre de Responsabilidad Ultima. El libertarismo parte de una tesis familiar: si el determinismo es verdadero, entonces cada decisión que adopte, igual que cada bocanada de aire que tome, es en último término un efecto de cadenas causales que se remontan a tiempos anteriores a mi nacimiento. En el capítulo anterior argumenté que la determinación no es lo mismo que la causación, es decir, que el conocimiento de que un sistema es determinista no nos dice nada sobre las relaciones causales interesantes —o de la *ausencia* de relaciones causales— entre los eventos que se producen en él; pero se trata de una conclusión controvertida, que se enfrenta a una larga tradición. Algunos pueden verla, en el mejor de los casos, como una excéntrica recomendación acerca de cómo usar la palabra «causa», de modo que será mejor dejarla a un lado temporalmente y ver qué ocurre si nos mantenemos aferrados a la tradición y tratamos el determinismo como la tesis de que cada estado de cosas *causa* el estado siguiente. En este caso, y tal como muchos han sostenido, si mis decisiones vienen causadas por cadenas de eventos que se remontan a tiempos anteriores a mi nacimiento, soy tan responsable *causalmente* de los resultados de mis acciones como pueda serlo un árbol al que se le cae una rama en plena tormenta de la muerte de la persona que se encuentra debajo; sin embargo, no es *culpa* de la rama que no fuera más resistente de lo que era, o que el viento soplara con tanta furia, o que el árbol creciera tan cerca del camino. Para ser moralmente responsable, debo ser la fuente última de mi decisión, y eso sólo puede ser cierto si las influencias previas no eran *suficientes* para garantizar el resultado, que «dependía verdaderamente de mí». Harry Truman tenía un famoso cartel en su escritorio del Despacho Oval que decía: «La cadena termina aquí». Una mente humana tiene que ser un lugar donde termina la cadena, dice Kane, y sólo el libertarismo puede

1. Seguido por una respuesta a sus críticos en «Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism» (1999).

aportar esta clase de libre albedrío, el único que puede darnos la Responsabilidad Ultima. Una mente es un terreno de «actos de voluntad (elecciones, decisiones o iniciativas)» y:

> Si dichos actos de voluntad estuvieran a su vez causados por otra cosa, de modo que las cadenas explicativas pudieran remontarse a la herencia o al entorno, a Dios o al destino, entonces la ultimidad no estaría en los agentes, sino en otra cosa (Kane, 1996, pág. 4).

Los libertaristas tienen la necesidad de encontrar una manera de romper esas odiosas cadenas causales en el momento en que el agente toma su decisión, pero, tal como reconoce Kane, el inventario de modelos libertaristas propuestos hasta el momento es un zoo de monstruos inservibles. «Los libertaristas han invocado centros de poder transempíricos, egos inmateriales, yoes nouménicos, causas inobjetivables, y toda una letanía de otras instancias especiales cuyas operaciones no quedaban muy claramente explicadas» (pág. 11). La intención de Kane es corregir esta deficiencia.

Antes de pasar a examinar su propuesta, sin embargo, deberíamos señalar que algunos libertaristas no ven aquí ninguna deficiencia. Algunos dualistas impenitentes, entre otros, abrazan la idea de que la existencia de la libertad requiere algo así como un milagro. Tienen la certeza íntima de que la libertad, la verdadera libertad, es estrictamente imposible en un mundo materialista, mecanicista y «reduccionista». ¡Tanto peor para la visión materialista! Consideremos, por ejemplo, la doctrina conocida como «causación por el agente». Roderick Chisholm, el principal arquitecto de la versión contemporánea de esta antigua idea, la define del siguiente modo:

> Si somos responsables [...] gozamos de una prerrogativa que algunos atribuirían únicamente a Dios: cada uno de nosotros, al actuar, es un primer motor inmóvil. Al hacer lo que hacemos, causamos ciertos eventos, y nada —o nadie— es causa de que nosotros causemos dichos eventos (Chisholm, 1964, pág. 32).

¿Cómo podemos causar «nosotros» estos eventos? ¿Cómo puede un agente causar un efecto sin que haya un evento (en el agente, en principio) que sea la causa de tal efecto (y que sea en sí misma el efecto de una causa anterior, y así sucesivamente)? La causación por el agente es una doctrina francamente misteriosa, cuya tesis no encuentra paralelo en nada de lo que descubrimos en los procesos causales de las reacciones químicas, la fisión o

la fusión nuclear, la atracción magnética, los huracanes, los volcanes, o procesos biológicos como el metabolismo, el crecimiento, las reacciones inmunológicas y la fotosíntesis. ¿Existe algo así? Cuando los libertaristas insisten en que es preciso que exista, caen en la trampa de los que están en el otro extremo, los deterministas duros, que no tienen problema en dejar que la insobornable definición de la libertad de los libertaristas siente los términos del debate, para que ellos puedan declarar a su vez, con la ciencia como aliada, que tanto peor para la libertad. Según he podido comprobar, aquellos que consideran evidente el carácter ilusorio de la libertad tienden a buscar su definición de la libertad en las formulaciones de los más radicales defensores de la causación por el agente.

Esta polarización es probablemente inevitable. Cuando hay mucho en juego, es aconsejable ser prudente, aunque un exceso de prudencia lleva a un encastillamiento de las posiciones y a la paranoia por la posible «erosión». Tal como se dice, si no eres parte de la solución eres parte del problema. Cuidado con cruzar la línea, con caer por la pendiente. Dales la mano y se tomarán el brazo entero. Sin embargo, la prudencia también puede llevar a que uno se convierta inconscientemente en su propia caricatura. En su celo por proteger algo que considera valioso, la gente a veces opta por marcar la línea demasiado lejos, convencida de que es más seguro defender demasiado territorio que demasiado poco. El resultado es que terminan por defender lo indefendible, y se aferran a una posición extrema que resulta vulnerable precisamente por su exageración. En todo caso, el absolutismo es una deformación profesional en filosofía, ya que las posiciones radicales y extremas son más fáciles de definir claramente, son más memorables y tienden a atraer más atención. Nadie se ha hecho nunca famoso como filósofo por promover el hibridismo ecuménico. En el tema de la libertad esta tendencia se ve amplificada y sostenida por la propia tradición: tal como han dicho los filósofos desde hace dos milenios, o somos libres o no lo somos; es todo o nada. Y así es como las diversas propuestas de compromiso, las sugerencias de que el determinismo pueda ser compatible con al menos *algunos* tipos de libertad, encuentran resistencia como malos negocios, peligrosas subversiones de nuestros fundamentos morales.

Los libertaristas llevan tiempo insistiendo en que las concepciones *compatibilistas* de la libertad como la que defiendo aquí no son realmente lo que buscamos, ni siquiera un sustituto aceptable, sino más bien un «miserable subterfugio», según la muy citada frase de Immanuel Kant. No hacen falta más que dos para jugar a este juego de descalificaciones. Observen ustedes mismos. A nosotros, los compatibilistas, nos parece que los

libertaristas ponen como condición de la libertad que se pueda realizar lo que podríamos llamar la *levitación moral.* ¿No sería maravilloso que fuéramos capaces de levitar, y luego lanzarnos en cualquier dirección siguiendo nuestro capricho? Me encantaría hacer eso, pero no puedo. Es imposible. No hay tales entes milagrosos como los levitadores, pero sí hay algunos casi levitadores bastante buenos: por ejemplo los colibríes, los helicópteros, los dirigibles y los planeadores. A los libertaristas, sin embargo, no les parece que la casi levitación sea suficiente, y consideran, en efecto, que:

> Si tienes los pies en el suelo, la decisión no es realmente tuya: es el planeta Tierra quien la toma en realidad. La decisión no *la has tomado* tú, sino que es la suma de las líneas causales que se cruzan en tu cuerpo, que no es sino un bulto móvil sobre la superficie del planeta, azotado por toda clase de influencias, sometido a la gravedad. La verdadera autonomía, la verdadera libertad, requiere que quien toma la decisión esté de algún modo suspendido, aislado del tira y afloja de aquellas causas, de modo que cuando se tomen las decisiones la única causa de ellas seas *¡tú!*

Éstas son las caricaturas. Tienen su utilidad, pero pongámonos serios y consideremos el intrépido intento de Kane de llenar las lagunas y ofrecer un modelo libertarista sobre la capacidad de tomar decisiones responsables. Tras reconocer que «la *libertad* es un término que tiene muchos significados», Kane concede que «*incluso aunque viviéramos en un mundo determinista,* podríamos distinguir con sentido entre personas que están libres de cargas tales como limitaciones físicas, adicciones o neurosis, coerciones u opresiones, y personas que no están libres de estas cargas, y podemos conceder que incluso en un mundo determinista merecería la pena preferir estas libertades a sus opuestos» (Kane, 1996, pág. 15). Así pues, algunas libertades que merecen la pena son compatibles con el determinismo, pero «las aspiraciones humanas trascienden» dichas libertades; «hay *al menos un* tipo de libertad que es incompatible con el determinismo, y es un *tipo significativo de libertad que merece ser deseada».* Es «el poder de ser el creador y el valedor último de los propios fines y propósitos» (pág. 15).

Se supone comúnmente que en un mundo determinista no hay *verdaderas* opciones, sino sólo opciones aparentes. En los dos capítulos anteriores he demostrado que esto es una ilusión, pero por más que lo sea también es cierto que se trata de una ilusión particularmente persistente y tentadora. Si el determinismo es verdadero, entonces en cada instante hay un único futuro físicamente posible, de modo que toda elección ha sido ya determinada, toda la vida no es otra cosa que el desarrollo de un guión que estaba

escrito desde el origen de los tiempos. Sin verdaderas opciones, sin encrucijadas en la propia trayectoria a lo largo de la historia, parece que difícilmente podemos considerarnos los *autores* de nuestros propios actos; somos más bien como actores de teatro, que recitan las líneas con aparente convicción, y cometen sus «crímenes» con gracia o torpeza, según las directrices que le hayan marcado. Convincente, ¿verdad? Y sin embargo, es falso. Probablemente la mejor manera de hacer evidente la sorprendente conclusión de que esta visión es simplemente falsa —una reacción de pánico que no queda justificada por la premisa del determinismo— es darle a la otra parte la oportunidad de decir qué es lo que nos *daría* verdaderas opciones. El reto al que se enfrenta Kane es ofrecer un modelo de acuerdo con el cual nuestras decisiones *aparentes* pudieran ser *verdaderas* decisiones, y pretende hacerlo además sin postular ninguna entidad sobrenatural ni misteriosas formas de agencia. Kane es naturalista, igual que yo, y parte de la base de que somos criaturas pertenecientes al orden natural cuya actividad mental depende de las operaciones de nuestro cerebro. Este requisito del naturalismo suscita algunas cuestiones que merece la pena examinar. (En capítulos posteriores estudiaremos más de cerca lo que tienen que decir la neurociencia y la psicología cognitiva contemporáneas acerca de la toma de decisiones, para ver si obtenemos algún resultado interesante cuando nos volvemos más ambiciosos y tratamos de agregar más detalles al cuadro.)

¿DÓNDE DEBEMOS PONER EL TAN NECESARIO VACÍO?

Una legendaria reseña literaria comienza diciendo: «Este libro viene a llenar un muy necesario vacío», y, lo dijera o no en serio el autor de la reseña, no hay duda de que Kane necesita ciertamente un vacío, un hiato en el determinismo, y pretende instalarlo en lo que llama la *facultad de la razón práctica* del cerebro. Describe esta facultad en términos de input, output," y lo que a veces ocurre durante el proceso intermedio entre ambos (véase la figura 4.1). Kane distingue estos tres fenómenos en términos de tres sentidos de la *voluntad-*.

*** Utilizo los términos originales por estar muy extendido su uso en nuestro contexto lingüístico y porque sus equivalentes en castellano (entrada y salida de datos) tienden a concretar demasiado el sentido de la expresión. (N. *del t.)***

(i) *voluntad, desiderativa* o *apetitiva:* lo que *quiero, deseo* o *prefiero* hacer
(ii) *voluntad racional*: lo que *escojo, decido* o *tengo intención* de hacer
(iii) *voluntad de perseverar [striving will*\: lo que *intento*, pongo mi *empeño* o mi *esfuerzo* en hacer (Kane, 1996, pág. 26).

A grandes rasgos, las voluntades del tipo (i) constituyen el input para la facultad de la razón práctica, la cual da como output voluntades del tipo (ii), cuando todo va bien. Cuando se produce una tensión dentro de la maquinaria lo que obtenemos es (iii), que siempre implica una resistencia y, por lo tanto, una pugna y un mayor esfuerzo. Todo esto suena bastante correcto y familiar. Cuando estamos indecisos, cargamos nuestra mente con cualesquiera preferencias o deseos relevantes que se nos ocurren (i), nos recordamos a nosotros mismos los datos y las creencias relevantes, y nos ponemos a reflexionar. Nuestras reflexiones, sean fáciles o penosas (iii), llevan finalmente a decisiones (ii). «Si hay alguna indeterminación en el libre albedrío, ésta debe encontrarse, en mi opinión, en algún lugar entre el input y el output» (Kane, 1996, pág. 27).

Kane propone un ejemplo para que podamos ver este sistema en funcionamiento: consideremos el caso de una mujer de negocios «que está de

FIGURA 4.1. Facultad del razonamiento práctico.

camino hacia una reunión importante para su carrera, cuando observa un asalto en una calle. A continuación se produce un conflicto interior entre su conciencia moral, que le reclama detenerse y pedir ayuda, y sus ambiciones profesionales, que le dicen que no puede faltar a la reunión» (Kane, 1996, pág. 126). Kane aventura la idea de que este conflicto podría activar dos «redes neurales recurrentes y conectadas», una para cada aspecto de la cuestión. Estas dos redes interconectadas se retroalimentan mutuamente, interactúan de múltiples maneras, interfieren la una con la otra y, en general, siguen revolviéndose hasta que una de ellas consigue tirar más fuerte de la cuerda, momento en que el sistema recupera el equilibrio y da una decisión como output.

> Esta clase de redes ponen en circulación impulsos e información en bucles que se retroalimentan y en general desempeñan un papel en el complejo procesamiento cognitivo que uno esperaría que tuviera lugar dentro del cerebro en el caso de la deliberación humana. Es más, las redes recurrentes son no lineales, lo que permite (tal como sugieren ciertos estudios recientes) la posibilidad de una *actividad caótica* [la cursiva es mía], lo que contribuiría a la plasticidad y la flexibilidad que manifiestan los cerebros humanos en la resolución creativa de problemas (de la que la deliberación práctica es un ejemplo). El input de una de esas redes recurrentes consiste en las motivaciones morales de la mujer, y el output en la decisión de volver atrás; el input de la otra, en sus ambiciones profesionales, y su output, en la decisión de seguir adelante hacia la reunión. Las dos redes están conectadas, de modo que el *indeterminismo que creaba la incertidumbre* [la cursiva es mía] respecto a si tomaba la decisión moral provenía de su deseo de hacer lo contrario, y viceversa, con lo que el determinismo surge, tal como hemos dicho, de un conflicto en la voluntad (Kane, 1999, págs. 225-226).

Antes de ir más lejos, debemos separar dos cuestiones que aparecen mezcladas en este pasaje. La «actividad caótica» que menciona Kane es un caos *determinista,* la impredecibilidad *práctica* de cierta clase de fenómenos que pueden describirse perfectamente dentro de la vieja física newtoniana. Tal como reconoce Kane, las dos redes que interactúan caóticamente no crearían en sí mismas ningún indeterminismo, de modo que si hubiera algún «indeterminismo que creara incertidumbre» tendría que proceder de otra parte. Este es un punto crucial. Kane no es el único en ver la importancia del caos en la toma de decisiones, pero su idea consiste en *suplementar* el caos con una pizca de aleatoriedad cuántica, siguiendo como muchos otros la estela de Roger Penrose (1989, 1994). La cuestión que debemos considerar es

si el ingrediente extra de Kane aporta algo importante, y para ello necesitamos tener una idea más clara de en qué consiste un fenómeno caótico.

Consideremos la pieza *Hyatt New Departure Ball Bearing*. Durante muchos años, el Museo de Ciencia y Tecnología de Chicago tenía expuesta una vitrina de cristal donde tenía lugar un fenómeno asombroso, hora tras hora. Esta pieza, donada por una rama de la General Motors, mostraba una interminable procesión de pequeñas bolas de acero que salían rodando de un pequeño agujero situado en la parte de atrás de la vitrina, caían varios pies hasta la muy pulimentada superficie de un «yunque» cilíndrico magníficamente bruñido, saltaban por el aire para pasar por el centro de un anillo que rotaba como una moneda sobre una mesa (de modo que el ritmo de los saltos a través del anillo en rotación tenía que ser exquisitamente preciso) y luego saltaban desde un segundo yunque hasta un pequeño agujero en la parte de atrás de la vitrina por el que hacían un medido mutis: salto, salto, susurro, salto, salto, susurro, cientos de veces cada hora. El cartel que había encima decía: «Esta máquina demuestra la precisión de la manufactura y la uniformidad de las propiedades físicas de las bolas empleadas en los cojinetes de bolas». En cuanto los dos yunques estaban debidamente ajustados, podía seguir funcionando días y días, durante los cuales cada bola seguía exactamente la misma trayectoria que su predecesora, un movimiento perfectamente predecible, fiable, determinista, una convincente demostración de que las propiedades físicas pueden fijar el propio destino (al menos si uno es una pequeña bola de metal). Su predictibilidad habría desaparecido por entero, sin embargo, si simplemente hubiéramos duplicado el número de yunques (de modo que cada bola tuviera que dar cuatro saltos antes de salir) y si hubiéramos puesto los yunques de lado, de modo que las bolas tuvieran que saltar por sus lados redondeados en lugar de por sus extraordinariamente lisas superficies superiores. Los márgenes de error en el pulido de las bolas y el ajuste de los yunques se acercarían peligrosamente a cero.[2] ¡La mera presencia de observadores al otro lado del cristal crearía la suficiente interferencia *gravitacional* como para alterar los cálculos más precisos y hacer que muchas de las bolas no llegaran a su destino final!

Esta clase de caos es determinista, pero no por ello deja de ser interesante; ciertamente podría, tal como dice Kane, «contribuir a la plasticidad y

2. El físico Michael Berry (1978) ha hecho los cálculos para predecir la trayectoria de las bolas de acero que rebotan en los topes redondos de las máquinas del millón. Tres rebotes nos llevan más allá de los límites de los cálculos factibles.

la flexibilidad que manifiestan los cerebros humanos». En años recientes se han estudiado y demostrado a través de muchos modelos —a los que alude Kane— las potencialidades de este tipo de caos y, en general, de la no-linealidad. Parte de esta investigación ha sido saludada por los críticos como el toque de difuntos para la Inteligencia Artificial (IA) o, más específicamente, para la variedad de la misma conocida como GOFAI —Good Oíd Fashioned Artificial Intelligence [«La vieja y desfasada inteligencia artificial»] (Haugeland, 1985)—, y en muchos círculos está cada vez más extendida la impresión de que las redes neurales no-lineales tienen potencialidades asombrosas que superan con mucho las de los simples ordenadores, con sus torpes y frágiles programas algorítmicos. Pero lo que han pasado por alto muchos fans de las redes neurales es el hecho de que los modelos mismos hacia los que señalan para demostrar sus tesis son modelos *informáticos,*, no sólo estrictamente deterministas, sino también algorítmicos cuando bajamos a su sala de máquinas. Sólo dejan de ser algorítmicos *en el nivel más alto.* (¿Puede el todo ser más «libre» que las partes? Tal vez sea ésta una de las maneras de conseguirlo.) Incluso un comentarista tan astuto como Paul Churchland puede caer en esta trampa tan tentadora. Al criticar, con toda la razón, el intento de Roger Penrose de alistar a la física cuántica en su campaña contra los terribles algoritmos de la IA, Churchland escribe:

> *No hace falta ir tan lejos como al reino de la física cuántica para encontrar un rico dominio de procesos no algorítmicos.* Los procesos que tienen lugar en una red neural de *hardware* [la cursiva es mía] son típicamente no algorítmicos, y constituyen la base de la actividad computacional que tiene lugar en nuestras cabezas. Son no algorítmicos en el sentido básico de que no consisten en una serie de estados físicos discretos que se suceden serialmente siguiendo las instrucciones de un conjunto de reglas de manipulación de símbolos previamente almacenadas (Paul Churchland, 1995, págs. 247-48).

Nótese la inclusión de la palabra «hardware». Sin ella, lo que dice Churchland sería falso. En realidad, todos los resultados a los que alude Churchland (las NETTalk, las redes capaces de aprender gramática de Elman, el EMPATH de Cottrell y Metcalfe, y otros) no son el resultado de «redes neurales de hardware», sino redes neurales virtuales simuladas en ordenadores estándar. Y, por lo tanto, a nivel básico, cada una de esas demostraciones *sí* «consistió en una serie de estados físicos discretos que se suceden serialmente siguiendo las instrucciones de un conjunto de reglas de manipulación de símbolos previamente almacenadas». Sin duda

no es éste el nivel que explica su peculiar potencial, pero es en todo caso un nivel algorítmico. Nada de lo que hacen esos programas trasciende los límites de la computabilidad de Turing. Del mismo modo que tuvimos que ir al nivel del juego de ajedrez para explicar la diferencia de capacidades entre los programas A y B en el capítulo 3, tenemos que ir al nivel del diseño de la red neural para explicar las notables virtudes de dichas redes simuladas, pero en ambos casos lo que ocurre al nivel micro es un proceso determinista, digital, algorítmico. Los propios modelos que trata Churchland en términos tan favorables se implementan como programas informáticos: algoritmos, desde el punto de vista de los límites de computabilidad. Así pues, a menos que Churchland pretenda desautorizar sus propios ejemplos favoritos, debe conceder, después de todo, que los procesos algorítmicos pueden exhibir las potencialidades que considera cruciales para la explicación de la mentalidad. Pero en tal caso, aunque fuera cierta su tesis de que las redes neurales de *hardware* no son algorítmicas, eso no serviría para explicar su potencial (puesto que las aproximaciones algorítmicas de las mismas poseen todas las virtualidades necesarias).[3]

Tanto el sencillo mundo Vida tratado en el capítulo 2 como los programas de ajedrez tratados en el capítulo 3 son digitales y deterministas, y lo mismo sucede, a pesar de todas sus potencialidades suplementarias, con las simulaciones informáticas de redes neurales no lineales. El ingrediente extra de Churchland —el hardware en lugar del software virtual— no añade nada al potencial de las redes neurales. O, si lo hace, nadie nos ha dado ninguna razón para que pensemos que es así.[4] ¿Añade algo el ingrediente extra de Kane (el indeterminismo del nivel cuántico)? Para responder a esta pregunta, debemos examinar los detalles. ¿Dónde y cómo pretende Kane insertar el indeterminismo que busca conseguir?

3. Este párrafo está tomado, con revisiones, de Densmore y Dennett, 1999.

4. Podría haber una razón, implícita en mi análisis del capítulo 2 sobre el papel de los encuentros fortuitos para la creatividad. Es posible que ninguna simulación informática factible, ningún mundo virtual lo bastante reducido como para ser simulable, pueda tener la mezcla de ruido y silencio necesarias para generar una capacidad creativa abierta. Eso no sería relevante para la tesis de Churchland sobre las redes neurales, pero podría ser verdad. El trabajo de Adrián Thompson (por ejemplo, Thompson y otros, 1999) sobre la electrónica evolutiva sugiere desde un campo distinto que el software no siempre puede sustituir al hardware en la exploración de las posibilidades de diseño. Thompson ha creado chips de hardware con potencialidades que no dependen de sus capacidades a nivel de software, sino de interacciones en el nivel microfísico que no son fruto del diseño y pueden ser seleccionadas por evolución artificial.

EL MODELO INDETERMINISTA DE KANE PARA LA TOMA DE DECISIONES

¿Qué debería hacer la facultad de razonamiento práctico, y cómo debería hacerlo? ¿Cuáles son las especificaciones, tal como diría un ingeniero, de su dispositivo de toma de decisiones? Kane dice que debería discernir de algún modo el peso relativo de las diversas razones y preferencias que se le presentan, y optar por la razón «por la que quiere actuar el agente, más de lo que quiere hacerlo por cualquier otra razón (para actuar de otro modo)». Añade la condición ulterior de que los casos de ejercicio fructífero o exitoso de la facultad no deberían ser el resultado de la coerción o la compulsión (Kane, 1996, pág. 30). Kane deja abierta deliberadamente desde el principio la cuestión de si la facultad opera de manera determinista, ya que, de acuerdo con su tesis, para que esta facultad dé como resultado una voluntad libre en el sentido libertarista, se requiere el ingrediente extra del indeterminismo. Al considerar las especificaciones para una facultad de la razón práctica, resulta útil ir más allá de las condiciones mínimas de Kane y considerar algunos de los tipos de *incompetencia* que no desearíamos que exhibiera nuestra facultad:

1. No da ningún output: está simplemente averiada. Somos incapaces de pensar lo que vamos a hacer a continuación.
2. Tiene una banda demasiado estrecha (no puede procesar simultáneamente todas nuestras querencias, deseos o preferencias, y se cuelga por incapacidad de digerir su inmenso input).
3. Produce el output con demasiada lentitud para el mundo en que vivimos.
4. Tiene el problema de Hamlet (bucle infinito) y retrasa indefinidamente su output.
5. Falla sistemáticamente a favor de ciertas *clases* de input (consejos de mamá, consideraciones relacionadas con el patriotismo, el sexo, la propiedad...).
6. Da un resultado *incorrecto* en relación con el input (por ejemplo, preferimos claramente los derechos humanos a tomar un helado en el tiempo *t,* pero nuestra facultad hace que *decidamos* comprar un helado en lugar de dar nuestro dinero a Amnistía Internacional).

Esta última idea plantea una interesante cuestión en relación con la debilidad de la voluntad, y con la voluntad de perseverar —el tipo (iii) de Kane— que surge cuando hay una resistencia y algo tiene que ceder.

¿Dónde está el *embrague* de este mecanismo? ¿Está dentro o fuera de la facultad?

El ejemplo ofrecido en (6) sitúa el embrague en el interior de la facultad, lo que permite desfases no deseados entre el input y el output: llegamos a una decisión no deseada. Pero parece haber un segundo caso posible: nuestro razonamiento práctico funciona perfectamente, de modo que decidimos gastar el dinero en derechos humanos, pero (lástima) el embrague se activa después de tomar la decisión y terminamos por comprar el helado en lugar de hacer lo que habíamos decidido hacer. (Véase la figura 4.2.) ¿Son realmente casos distintos? Si es así, ¿cuál es la diferencia, y qué importancia tiene? ¿En qué punto se convierte la decisión en una auténtica decisión? Éste no es el único problema de límites que encontraremos.

¿Qué pasaría si nuestra facultad de razonamiento práctico diera outputs distintos para los mismos inputs? ¿Sería esto un defecto? Normalmente queremos que los sistemas sean fiables, y con esto nos referimos a que podamos contar con que den siempre el mismo output —el *mejor*

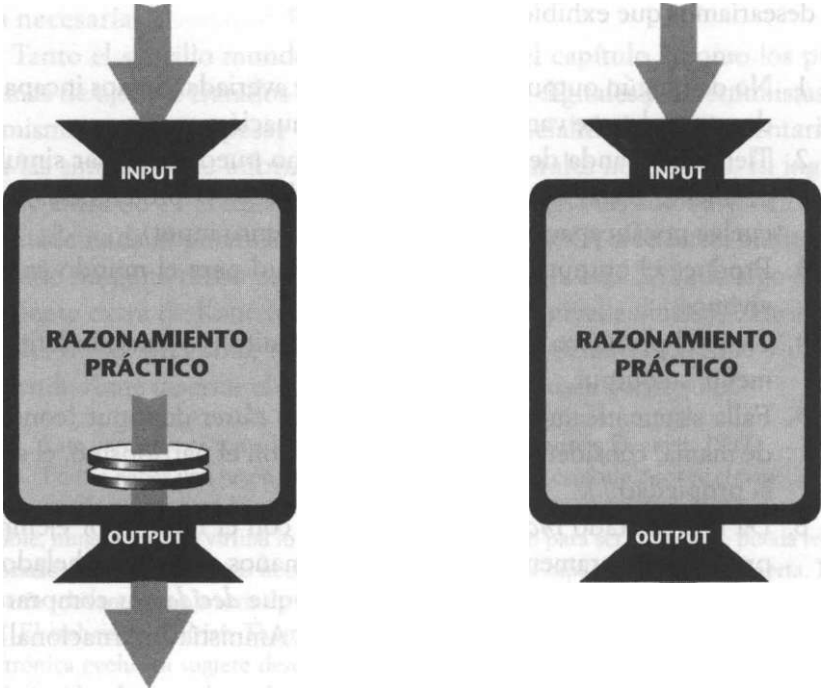

FIGURA **4.2. Posiciones del embrague, dentro y fuera.**

output, sea el que sea— para cada input posible. Tomemos como ejemplo nuestra calculadora. A veces, cuando no es posible definir cuál es el mejor output o cuando queremos específicamente que el sistema introduzca una variación «aleatoria» en el supersistema general, nos parece bien que dé diferentes outputs para el mismo input. La forma habitual de conseguirlo consiste en incorporar un generador de números pseudoaleatorios en el sistema que haga las veces de una moneda lanzada al aire (y genere un 0 o un 1 cada vez que se le pida), de un dado normal de seis caras (y genere un número entre 1 y 6 cada vez que se le pida) o de una rueda de la fortuna (y genere un número entre 1 y *n* cada vez que se le pida). Kane quiere algo mejor que la pseudoaleatoriedad. Quiere aleatoriedad genuina, y se propone conseguirla suponiendo la existencia de algún tipo de amplificador de las fluctuaciones cuánticas en las neuronas. Tal como vimos en el capítulo anterior, esto no haría que su modelo fuera más flexible o abierto, más capaz de mejorar o de aprender. No le daría a su sistema ninguna oportunidad que no tuviera con un generador de números pseudoaleatorios, pero no es ésa su función. Su función es metafísica, no práctica.

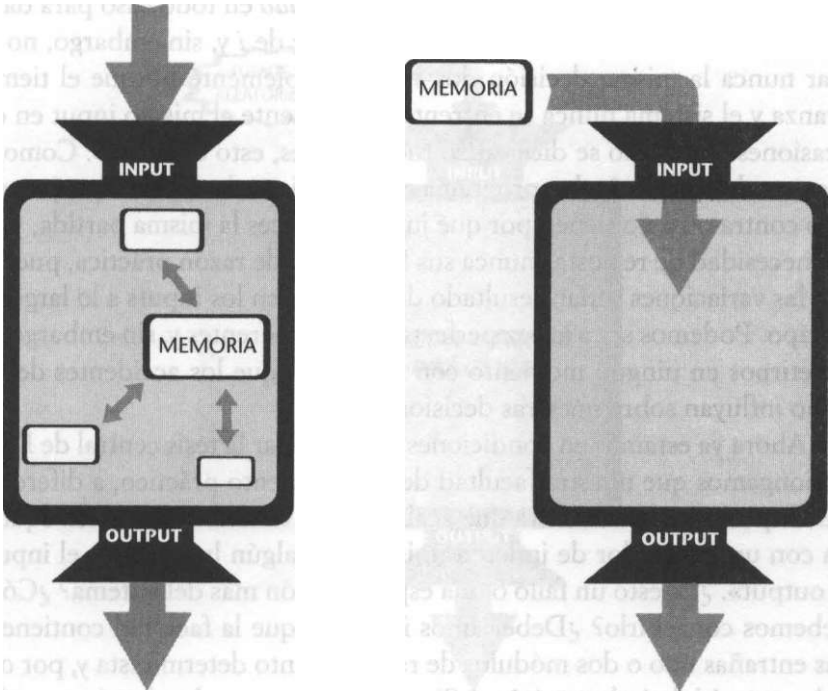

FIGURA 4.3. **Posiciones de la memoria, dentro y fuera.**

En todo caso, ¿*deberíamos* querer que nuestra facultad de razonamiento práctico diera diferentes outputs para los mismos inputs? Nos enfrentamos aquí a otro problema de límites. ¿Qué consideramos un input? ¿Contiene dicha facultad la historia de sus actividades previas, o es sólo la maquinaria vacía de todo contenido, el procesador, que recibe (parte de) dicha historia desde una memoria externa? (Véase la figura 4.3.)

No nos gustaría que nuestro razonamiento práctico fuera tan rígido que tomara la misma decisión cada día: por ejemplo, decidirse siempre por un bocadillo de jamón para el almuerzo. Pero si incluimos en los inputs posibles hechos procedentes de la memoria de modo que uno de los inputs de hoy sea el hecho de que llevamos dos días comiendo bocadillo de jamón, ello haría que el caso de *hoy* fuera distinto del de ayer, con independencia de cuál sea la decisión final. Como las personas poseen una memoria y una capacidad perceptiva notables, nunca se encuentran dos veces en el mismo estado exacto, de modo que pueden obtener gran variabilidad en el output de sus facultades de razón práctica por el mero hecho de alimentarla con un input más variado acerca de su estado y sus circunstancias actuales. Nuestro sistema de razonamiento práctico debería ser tan fiable como una calculadora, *determinado* en todo caso para dar el output! en respuesta al input, para cada valor de *i* y, sin embargo, no tomar nunca la misma decisión dos veces, simplemente porque el tiempo avanza y el sistema nunca se enfrenta exactamente al mismo input en dos ocasiones. Tal como se dice: «Eso fue entonces, esto es ahora». Como vimos en el capítulo 3, dos programas informáticos de ajedrez que jueguen uno contra otro no tienen por qué jugar dos veces la misma partida, y eso sin necesidad de reajustar nunca sus facultades de razón práctica, pues todas las variaciones serían resultado de cambios en los inputs a lo largo del tiempo. Podemos ser a la vez perfectamente coherentes y, sin embargo, no repetirnos en ningún momento con sólo dejar que los accidentes del camino influyan sobre nuestras decisiones.

Ahora ya estamos en condiciones de examinar la tesis central de Kane. Supongamos que nuestra facultad de razonamiento práctico, a diferencia del dispositivo determinista que acabamos de describir, estuviera equipada con un generador de indeterminismo «en algún lugar entre el input y el output». ¿Es esto un fallo o una especificación más del sistema? ¿Cómo debemos concebirlo? ¿Deberíamos imaginar que la facultad contiene en sus entrañas uno o dos módulos de razonamiento determinista y, por otro lado, un módulo indeterminista? Si ponemos el generador de números aleatorios fuera de la facultad (figura 4.4), los números aleatorios que genera-

rá deberían considerarse inputs respecto a la facultad, y ésta debería tratarlos como cualquier otro input; si es fiable, debería producir un resultado *determinado* en relación con dicho input. Si, en cambio, ponemos un generador de números aleatorios en el interior de la facultad, para que sea más libre en su procesamiento de los inputs, entonces los outputs de la facultad *no* vendrán determinados por sus inputs. Sin embargo, no hemos hecho más que situar la línea en un lugar funcional diferente.

Kane dice que la indeterminación debería estar «entre» el input y el output, pero podríamos muy bien preguntarnos por qué la indeterminación no podría *formar parte* del input. ¿Qué diferencia supondría eso? Le planteé esta pregunta a Kane (al discutir un borrador previo de este capítulo), y dio una interesante respuesta:

> **Hay un razón por la que debe estar entre el input y el output y no formar parte meramente del input. La razón es que todo cuanto suponemos que ocurre entre el input y el output corresponde a la actividad del agente (en la forma de razonamiento práctico que tendrá como resultado la elección). El input (en la forma de disposiciones, creencias, deseos y cosas por el estilo) no es algo que el agente controle aquí y ahora, aunque parte de ello pueda ser el**

FIGURA 4.4. Generador de aleatoriedad externo.

producto de razonamientos, actividades o elecciones realizados en momentos anteriores [...]. El indeterminismo meramente en el input no da lugar a una responsabilidad fuerte. Para recoger plenamente la noción libertarista de responsabilidad, el indeterminismo debe ser un *ingrediente* no sólo de lo que le «viene a la cabeza», sino de lo que el agente está haciendo realmente (razonar, tomar decisiones). Si los inputs son el resultado de nuestras acciones, todo bien, pero si simplemente son algo que nos ocurre, no es suficiente, aunque sea por azar (Kane, correspondencia personal).

Kane quiere que el indeterminismo sea el «resultado de nuestras acciones» y no algo que «simplemente ocurre». Esto es fácil de arreglar: que la facultad de razonamiento práctico *pida por encargo* algo de aleatoriedad cada vez que, en el curso de sus labores, encuentre algo que interpreta como un bloqueo de algún tipo, sea un imponderable o sea una metaelección sobre qué pensar o dónde dirigir la atención a continuación (figura 4.5).

De este modo, como la aleatoriedad será «reclamada» como resultado de las actividades específicas de la facultad, no será algo que haya caído del cielo sin que nadie lo pidiera. Es más, el uso al que se aplique la aleatoriedad requerida vendrá determinado por las actividades constructivas de la

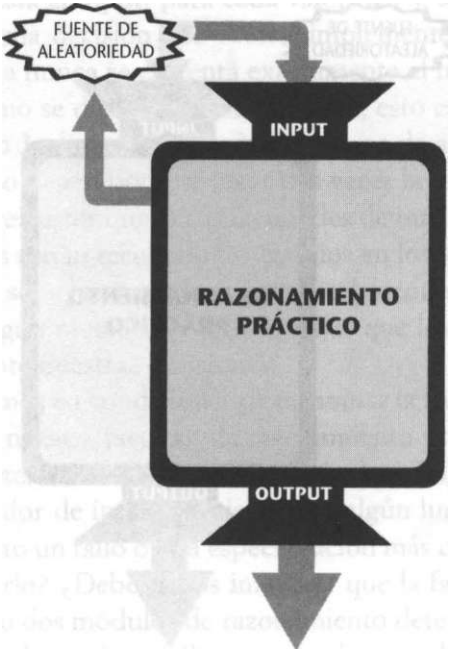

FIGURA 4.5. Aleatoriedad por encargo.

propia facultad. (Si decido lanzar una moneda para resolver dónde cenar hoy, sigue siendo una decisión mía; yo *hice* que eso resolviera la cuestión.) Pero una vez más no estamos haciendo otra cosa que marcar la línea en otro lugar; cualquier cosa que pueda conseguirse gracias a una fuente de aleatoriedad *interna* puede conseguirse también a base de su *importación* desde una fuente *externa* de aleatoriedad que pueda ser consultada cuando sea necesario. La metáfora del contenedor, por lo que llevamos visto hasta ahora, ocupa un lugar central en la teoría de Kane.

Sin embargo, por mor de la argumentación, supongamos que Kane puede darnos una buena razón para distinguir entre fuentes internas y externas de aleatoriedad. Instalamos la indeterminación en el interior de la facultad, entre el input y el output, de acuerdo con sus especificaciones, y luego instalamos la facultad en el interior del agente. ¿Cómo funciona en la vida cotidiana? Kane señala que

> las elecciones o decisiones son normalmente la culminación de procesos de deliberación o razonamiento práctico, pero no tienen por qué serlo siempre. No tenemos por qué descartar la posibilidad de que se tomen decisiones rápidas, impulsivas o irreflexivas, las cuales vienen a resolver también situaciones de indecisión, pero surgen de un razonamiento previo mínimo o nulo. Sin embargo, por más que puedan producirse decisiones rápidas o impulsivas de este tipo, son menos importantes para la libertad que las decisiones que culminan procesos de deliberación durante los cuales se consideran reflexivamente diversas alternativas. Ello es así porque en estos casos es más probable que nos parezca que controlamos el resultado y que «podríamos haber hecho otra cosa» (Kane, 1996, pág. 23).

De todo ello extraemos la imagen de unos actos *ocasionales* de elección deliberada que constituyen los puntos de inflexión moralmente relevantes —que «desempeñan un papel crucial» (pág. 24)— a partir de los cuales se establecen hábitos e intenciones sobre cuya base se actúa posteriormente de manera harto irreflexiva pero aún responsable. Consideremos, por ejemplo, una decisión rápida. Mi esposa me pregunta si puedo pasar por Correos de camino hacia el trabajo y enviar un paquete por ella, y le contesto casi instantáneamente que no puedo porque llegaría tarde a una cita con un alumno. ¿He deliberado? ¿Me embarqué en un proceso de razonamiento práctico? No se trata de una decisión moral de mucho calado, pero es el material del que están hechas en buena medida las vidas morales (e inmorales): cientos y miles de decisiones menores tomadas tras muy escasa consideración, habitualmente sobre el fondo de una justificación tácita e inarticulada. Sería

más bien extraño que respondiera algo del estilo: «Bien, como eres mi esposa y hemos prometido solemnemente ayudarnos el uno al otro, y como no encuentro ningún defecto o problema en tu requerimiento —no me has pedido que haga algo físicamente imposible, o ilegal, o autodestructivo, por ejemplo—, hay sin duda fuertes razones para decir "Sí, cariño". Por otro lado, le he dicho a un alumno que nos veríamos a las 9.30 y, considerando el tráfico, cumplir con tu requerimiento supondría tenerle esperando al menos media hora. Podría tratar de llamarle y pedirle permiso para cambiar la hora prevista, pero tal vez no lo encuentre, y, por otro lado, la cuestión importante es si enviar el paquete con tal presteza es una tarea lo bastante importante como para justificar que le cause un inconveniente al alumno. Convenir una cita con él equivale por mi parte a realizar una promesa, aunque de una clase que sería disculpable romper por causa...». Tal vez sorprenda la observación de que todas estas consideraciones (¡y muchas más!) realmente contribuyeron *en alguna medida* a mi respuesta. ¿Cómo es posible? Pues bien, ¿habría emitido yo un juicio rápido, positivo o negativo, si mi esposa me hubiera pedido que por favor estrangulara al dentista de camino al trabajo, o hiciera saltar el coche por un precipicio? Si lo que le hubiera dicho a mi alumno hubiera sido meramente que tenía previsto estar en mi oficina a las 9.30 para tomar café (sin hacer ni dar a entender ninguna promesa), o hubiera propuesto una hora más flexible para la entrevista, o hubiera estado hablando con él por teléfono en el momento mismo en que mi esposa hizo su petición, todo ello podría haber marcado una diferencia, sin duda, en mi juicio rápido. Incluso un juicio rápido puede ser notablemente sensible a una multitud de aspectos de mi mundo que han conspirado a lo largo del tiempo para crear mis disposiciones actuales.

Kane concede sin ningún problema que dicho complejo conjunto de disposiciones, que se ha venido gestando de manera más o menos continua desde que era un niño, pueda *determinar* mi respuesta en un caso como éste y en otros casos en los que no delibero. Pero una vez más acechan los problemas de límites. ¿Debemos considerar que los juicios rápidos *proceden de* la facultad de deliberación (pero de forma tan instantánea y directa que los detalles se mantienen tácitos) o deberíamos considerar que proceden más inmediatamente de algún subsistema o facultad «inferior», mientras que la facultad de deliberación se mantiene en reserva para casos de un calado especial? En mi opinión, lo mejor es establecer las demarcaciones (que al fin y al cabo no son sino líneas de análisis para los filósofos, no límites anatómicos por descubrir) de tal modo que también los juicios rápidos sean realizados, sin dificultad, en y por la facultad de

razonamiento práctico. Pues, aunque el paréntesis del indeterminismo debe localizarse en el interior de la facultad (entre el input y el output), Kane sostiene, tal como veremos, que la facultad no tiene por qué actuar siempre de manera indeterminista. En ciertas ocasiones puede operar de manera determinista, incluso al tratar cuestiones morales de peso. (¿Debo estrangular al dentista? ¡Anda ya!)

Kane no tiene mayor problema con este papel ocasional que desempeña el determinismo en la vida de un agente moral, y ello por diferentes motivos. En primer lugar, le permite dar una versión realista de dichos juicios rápidos. Simplemente no resulta plausible sostener qué los hábitos de toda una vida, que dan pie a decisiones tan predecibles que uno puede apostar la vida por ellas, sean sin embargo indeterministas (excepto en el sentido limitado de que pueda haber una posibilidad entre un cuatrillón de que puedan tener una excepción). Pensemos en nuestra disposición a conducir por la autopista, frente a coches que se acercan por el carril contrario a una velocidad superior a los 160 kilómetros por hora. Nuestra vida depende de que los conductores no decidan, como serían libres de decidir, pasarse repentinamente a nuestro lado de la autopista, sólo para ver lo que pasa. Nuestra ecuanimidad en la autopista demuestra hasta qué punto asumimos que es predecible el comportamiento de esos perfectos extraños. *Podrían* matarnos en un acto sin sentido, un *acte gratuit* suicida, pero no pagaríamos ni un dólar o un centavo de dólar por la oportunidad de vaciar la carretera de todos los coches que puedan venir en dirección contraria antes de salir. En segundo lugar, Kane necesita que el determinismo le eche una mano para dar respuesta a una objeción más seria contra el libertarismo que planteé yo mismo en *Elbow Rooms:* el caso de Martin Luther.

> «Aquí estoy —dijo Luther—. No puedo hacer otra cosa.» Luther afirmaba que no podía hacer otra cosa, que su conciencia le hacía *imposible* retractarse. Por supuesto, podría ser que estuviera equivocado o que estuviera exagerando deliberadamente la verdad. Pero, incluso aunque lo estuviera haciendo —tal vez sobre todo si era así—, su declaración es testimonio del hecho de que simplemente no eximimos a nadie de mérito o culpa por un acto porque pensemos que no podía hacer otra cosa. Fuera lo que fuera lo que estuviera haciendo Luther, no estaba tratando de escamotear su responsabilidad (Dennett, 1984, pág. 133).

Kane acepta que la decisión de Luther no podía estar más lejos del juicio irreflexivo, que era sin duda una decisión moralmente responsable, y que es muy posible que fuera cierto lo que dijo Luther respecto a ella: no podía

haber hecho otra cosa; verdaderamente estaba *determinado en el momento por su facultad de razonamiento práctico* para mantenerse firme. El caso de Luther no es un caso raro o carente de importancia. Tal como veremos en capítulos posteriores, la estrategia de prepararse a uno mismo para tomar decisiones difíciles a base de asegurarse de que uno esté determinado para hacer lo correcto cuando llegue el momento es uno de los umbrales de la responsabilidad madura, y Kane lo reconoce como tal. En realidad, Kane construye su teoría de la libertad alrededor de la idea de que todo agente moralmente responsable habrá pasado por una serie de ocasiones relativamente infrecuentes a lo largo de su vida en las que se habrá enfrentado a deseos contrapuestos, lo que da pie a una voluntad de perseverar del tipo (iii). En algunas de estas ocasiones habremos decidido realizar «acciones autoformativas» (AA), que pueden tener un efecto determinista sobre nuestro comportamiento subsiguiente; el único requisito es que dichas AA sean el resultado de procesos genuinamente indeterministas que hayan tenido lugar en el interior de la facultad de razón práctica:

> Un acto como el de Luther puede ser responsable en último término [...] aunque esté determinado por su voluntad, *porque* la voluntad de la que surgió es una voluntad *generada por él mismo*, y en este sentido es su «propia» voluntad [...]. Los actos responsables últimos, o los actos realizados por el propio libre albedrío, constituyen una clase de acciones que no se agota en las acciones autoformativas (AA) definidas por su indeterminación y porque el agente podría haber hecho otra cosa. Pero si no hubiera acciones «autoformativas» en este sentido, no conservaríamos la responsabilidad *última* por nada de lo que hiciéramos (Kane, 1996, pág. 78).

Cuando lanzo una piedra con una catapulta contra mi enemigo, la trayectoria de la piedra deja de estar en mis manos desde el momento en que se encuentra en el aire, y queda de este modo fuera del alcance de mi voluntad, pero los efectos de su caída siguen siendo responsabilidad mía, por mucho que tarde en producirse. Cuando me lanzo a mí mismo en una trayectoria de un cierto tipo, y me aseguro de que en lo sucesivo no podré modificar varios aspectos de esa trayectoria, la conclusión sigue siendo manifiestamente la misma. Reflexiones como ésta llevan a algunos libertaristas a aceptar que la libertad que tratan de implementar podría encontrarse concentrada en algunas ventanas de oportunidad con ciertas propiedades especiales. (Peter van Inwagen, por ejemplo, se suma a Kane en este punto, pero a diferencia de Kane supone que dichas ventanas deben de ser más bien infrecuentes.) Pero ¿de qué propiedades espe-

cíales estaríamos hablando? Kane dice que una AA debe cumplir la condición PA:

> (PA) El agente tiene *posibilidades alternativas* (o puede hacer otra cosa) respecto a A en t, en el sentido de que en *t* el agente *puede* (tiene el *poder* o la *capacidad* de) hacer A y *puede* (tiene el *poder* o la *capacidad* de) *hacer otra cosa* (Kane, 1996, pág. 33).

Nótese la función que cumple «en *t*» en esta fórmula. Algunos filósofos no pueden soportar decir cosas sencillas como: «Supongamos que un perro muerde a un hombre». Se sienten obligados a decir, en cambio: «Supongamos que el perro *p* muerde al hombre *h* en el tiempo *t*», con lo que demuestran su compromiso incorruptible con el rigor lógico, aunque no manejen luego ninguna fórmula que contenga *p,hy t.* Las referencias a *t* son omnipresentes en las definiciones filosóficas, aunque raramente desempeñan ningún papel relevante. Aquí, sin embargo, sí desempeña una función relevante. La definición se refiere a lo que es el caso en cada momento en el tiempo; requiere de nosotros que pensemos en «posibilidades en un determinado instante». Kane (pág. 87) cita un lírico pasaje de William James:

> La cuestión crucial [...] es que las posibilidades están realmente *aquí [...]* En aquellos momentos que ponen el alma a prueba, cuando la balanza del destino parece vacilar [...] [reconocemos] que la cuestión no se decide más que *aquí* y *ahora.* Eso es lo que da su realidad palpitante a nuestra vida moral y hace que nos estremezcamos [...] con una excitación extraña y sutil (James, 1897, pág. 183).

Examinemos más de cerca esta balanza que vacila. Imaginemos que nuestra facultad de la razón práctica lleva incorporado un dial con una aguja que marca la inclinación presente de la balanza en el curso de nuestras deliberaciones, una aguja que baila entre el Irse y el Quedarse (suponiendo que éstas sean las opciones que consideramos en este momento), cabecea del uno al otro y tal vez incluso vacila, en sus bruscas oscilaciones entre los dos valores (figura 4.6). Y supongamos que en cualquier momento podemos dar fin al proceso de deliberación presionando el botón *¡Ahora!,* con el que sellamos nuestra elección en favor de la opción —Irse o Quedarse— que resulta favorecida en aquel instante de la deliberación. Supongamos, de momento, que todo el procesamiento por parte de nuestra facultad del razonamiento práctico es determinista; se limita a «com-

parar pesos» de acuerdo con cierta función determinista aplicada a todos los inputs considerados hasta el momento, y da como resultado un valor constantemente actualizado que oscila de un lado al otro, entre el Irse y el Quedarse, en función del orden en el que son procesadas y reprocesadas las distintas consideraciones a la luz de las deliberaciones posteriores.

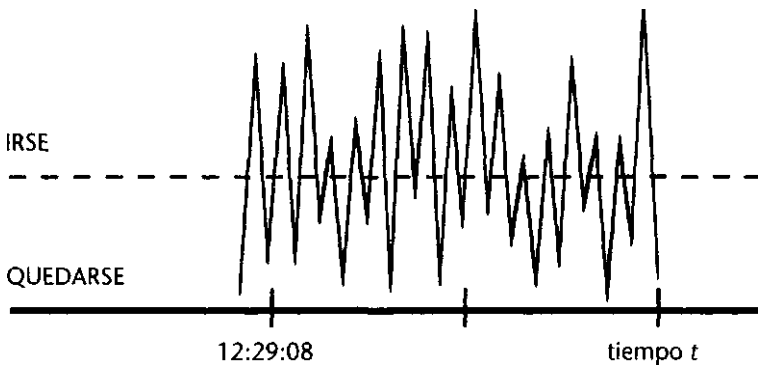

FIGURA 4.6. Aguja que oscila **entre** Irse **y** Quedarse.

¿Se cumpliría la condición PA en un caso como éste? ¿Dónde debemos mirar para encontrar la respuesta a esta pregunta? Supongamos que miramos al último minuto de la deliberación y observamos que en este período la aguja ha oscilado entre un lado y el otro doce veces o más, y que la aguja ha apuntado más o menos la mitad del tiempo hacia el Irse y la otra mitad hacia el Quedarse. Respecto a esta escala temporal parece claramente como si ambas alternativas estuvieran *abiertas* (en comparación, por ejemplo, con un minuto durante el cual la aguja apuntara firmemente hacia el Quedarse). Pero para Kane (y para James) no hay bastante con esto. Para que haya auténtica libertad, ambas posibilidades deben estar abiertas en el tiempo *t,* el instante mismo en el que se ha presionado el botón *¡Ahora!* Si nos concentramos en ese momento, y observamos que durante los 10 milisegundos previos al tiempo *t* la aguja apuntaba firmemente al Quedarse, que era también la decisión registrada en el momento de presionar el botón *¡Ahora!,* parecería que había buenos motivos para decir que la opción Irse no estaba disponible en el momento *t* (véase la figura 4.7).

¡Ah, pero hay una laguna! Hemos imaginado que habíamos presionado *nosotros* el botón *¡Ahora!* ¿Cómo podemos introducir la indeterminación si permitimos que el momento exacto en que se presiona el botón «dependa de nosotros»? Supongamos, pues, que el proceso de delibera-

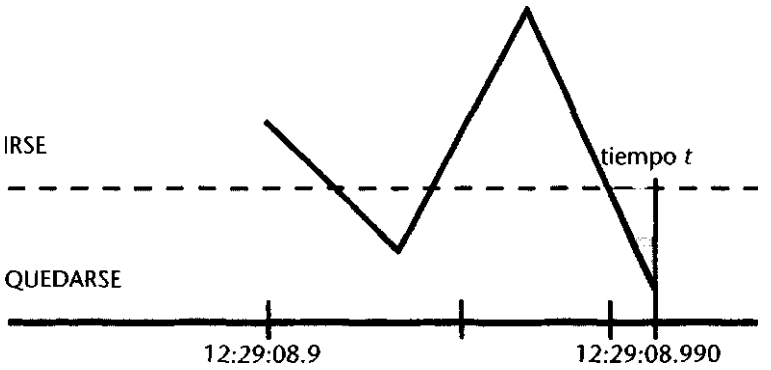

FIGURA 4.7. Ampliación de la figura 4.6 en el período de los 10 milisegundos.

ción está completamente determinado y que la indeterminación está en el momento exacto en que se presiona el botón *¡Ahora!* En algún momento de los próximos 20 milisegundos el botón será presionado, pero exactamente cuándo es algo que queda estrictamente (cuánticamente) indeterminado. En tal caso, si la oscilación entre el Irse y el Quedarse se produce con una frecuencia lo bastante alta como para que haya momentos de Irse y Quedarse en aquella ventana de 20 milisegundos, la decisión que se tome al activarse el botón *¡Ahora!* quedará indeterminada, completa y oficialmente impredecible desde una descripción completa del universo al comienzo de la ventana de oportunidad (figura **4.8).**

Por desgracia, seguirá sin cumplirse la condición PA, debido a un defecto en la definición de PA: esa fastidiosa cláusula «en *t*». Seguirá siendo

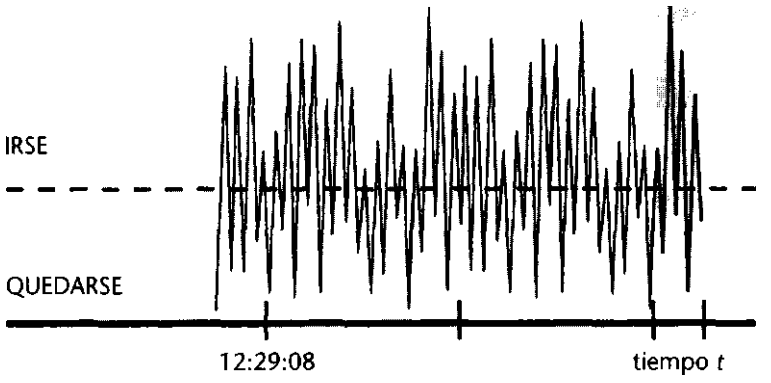

FIGURA 4.8. Ventana de oportunidad.

completamente predecible que si la decisión ocurre en el milisegundo 5, por ejemplo, será una decisión de Irse, y si ocurre en el milisegundo 17 será una decisión de Quedarse. En realidad, para cualquier tiempo *t* en la ventana de oportunidad está perfectamente determinada cuál será la decisión que se tome; lo que no está determinado es el momento exacto en que se tomará la decisión. El agente no es libre *en t* para Irse o Quedarse, para ningún valor de *t*. Pero ¿no podría haber suficiente con eso, mientras el instante de la elección siguiera siendo indeterminado? Resulta tentador proponer una revisión moderada de la condición PA que responda a nuestro sencillo modelo: que el tiempo *t* esté repartido por toda la ventana temporal de 20 milisegundos en lugar de concentrarse en un instante, en cuyo caso ya seríamos libres, puesto que tanto Irse como Quedarse coexisten en el tiempo *t* así extendido (y difícilmente puede considerarse que 20 milisegundos es un período largo de tiempo).

No hay duda de que la aguja en el dial y el botón hacen que este modelo tenga un aspecto terriblemente «mecánico», pero es el propio Kane quien lo quiere así. Pretende ser un libertarista *naturalista,* por lo que quiere que su modelo sea científicamente respetable, algo que pueda implementarse en el cerebro, y el dial y el botón son simplemente dispositivos que nos ayudan a visualizar el estado subyacente de complejidad neural relevante. Debe haber *algún* tipo de estado neural físicamente posible que realice la comparación de pesos, y *algún* tipo de transición entre estados que implemente la decisión (que produzca un output); podemos pretender simplemente que el dial se encarga de lo primero y el botón de lo segundo. De este modo el modelo ilustraría una manera posible —una familia de maneras— de amplificar la indeterminación cuántica subatómica para que desempeñe un papel crucial en la toma de decisiones. Es más, el modelo parece satisfacer el requisito de Ultimidad de Kane para las AA:

> (U) para cada X e Y (donde X e Y representan ocurrencias de eventos y/o estados) si el agente es personalmente responsable de X, y si Y es un *arche*[5] (es decir, una base, una causa o una explicación suficiente) para X, entonces el agente también debe ser personalmente responsable de Y (Kane, 1996, pág. 35).

Traducción: sólo podemos ser personalmente responsables de algo si somos personalmente responsables de todo cuanto sea condición suficiente para ello. Según Kane,

5. ***Arche*** **es el término aristotélico para referirse a** ***origen.***

> las AA son las acciones (u omisiones) voluntarias indeterminadas y últimas que se requieren en las historias vitales de los agentes para que se cumpla U (Kane, 1996, pág. 75).

La activación indeterminista del botón *¡Ahora!* convertiría la decisión en indeterminista en aquellos casos en los que ambas opciones oscilaran en una ventana de oportunidad ligeramente ampliada; en ningún momento anterior *habría* condición suficiente ni para la decisión de Irse ni para la de Quedarse, de modo que podríamos ser personalmente responsables de Irnos (o de Quedarnos) sin tener que preocuparnos por si somos responsables de alguna condición suficiente previa para Irse (o Quedarse). Por supuesto, todavía debemos encontrar alguna forma de dar sentido a que el accionamiento indeterminista del botón «dependa de *nosotros*» y no sea en sí mismo un input *externo* y aleatorio.

«SI UNO SE HACE LO BASTANTE PEQUEÑO, PUEDE EXTERNALIZARLO PRÁCTICAMENTE TODO»[6]

Una vez más nos encontramos con un problema de fronteras, y esta vez es serio: ¿cómo puede Kane conseguir que la indeterminación cuántica se quede en el *interior* del sistema relevante? Para ver dónde está la dificultad, supongamos que alguien que pasa por ahí grita justo cuando estamos a punto de presionar el botón *¡Ahora!,* y nos sorprende hasta el punto de hacernos presionar el botón cinco milisegundos antes, lo que lo convierte en la *causa* de que lo hayamos presionado. ¿Ya no es nuestra la decisión? Después de todo, la parte crucial de la relación causal, la parte que determinaba si debíamos Irnos o Quedarnos, vino en sí misma causada por el grito de ese alguien (que tenía su causa en una gaviota que volaba muy cerca, lo que a su vez fue causado por el regreso de la flota pesquera, lo que a su vez fue causado por el regreso de El Niño, lo que... vino causado por una mariposa que batió sus alas en 1926). Y aunque el batir de alas de esa mariposa estuviera verdaderamente indeterminado y no

6. Esta fue probablemente la frase más importante de *Elbow Room* (Dennett, 1984, pág. 143), y cometí el estúpido error de ponerla entre paréntesis. Desde entonces he tratado de corregir este error en mi obra, y de extraer las muchas implicaciones que tiene abandonar la idea de un yo puntual. Por supuesto, lo que pretendía subrayar con esta formulación irónica era la idea contraria: es sorprendente lo mucho que puede uno internalizar, si se hace lo bastante grande.

fuera sino el efecto magnificado de un salto cuántico en su minúsculo cerebro, este momento de indeterminismo se encuentra *en el lugar y el momento equivocados.* No será el momento de libertad de la mariposa en 1926 lo que nos haga libres a nosotros hoy, ¿verdad? El libertarismo de Kane requiere que rompa la cadena causal en algún lugar *dentro* del agente y *en* el momento de la toma de la decisión, el requisito del «aquí y ahora» del que tan elocuentemente hablaba William James. Si tan importante es, como sin duda piensan los libertaristas, será mejor que protejamos nuestros procesos de deliberación de todas esas interferencias externas. Haríamos mejor en aislar la pared que está alrededor de... *nosotros* mismos, para que las fuerzas externas no interfieran en la decisión que estamos preparando en nuestra cocina interna, sólo con los ingredientes que *nosotros* hemos permitido que entraran por la puerta.

Esta retirada del yo a un recinto vallado en el interior del cual se realiza todo el trabajo importante de creación tiene lugar en paralelo con otra retirada al centro del cerebro, siguiendo la desviada línea de argumentación o razonamiento que lleva a lo que llamo el Teatro Cartesiano, el lugar imaginario situado en el centro del cerebro donde «todo confluye» ante la conciencia. *No hay tal lugar,* y cualquier teoría que presuponga tácitamente que existe debe ser descartada de entrada por errónea. Las funciones que realiza el homúnculo imaginario en el Teatro Cartesiano deben ser distribuidas, en el cerebro, *en el tiempo y el espacio.* El problema que se le presenta a Kane es complejo, ya que debe encontrar alguna forma de conseguir que el evento cuántico indeterminado no sólo esté *en nosotros,* sino que también sea *nuestro.* Ante todo quiere que la decisión «dependa de nosotros», pero si la decisión es indeterminada —el requisito definitorio del libertarismo— tampoco puede estar determinada por nosotros, con independencia del tipo de ente que seamos, porque no está determinada por nada. Seamos lo que seamos, no podemos *influir* en un evento indeterminado —el sentido mismo de la indeterminación cuántica es que tales eventos cuánticos no están influidos por nada—, de manera que de algún modo tendremos que *convertirlo en nuestro socio* o *unir fuerzas* con él, ponerlo a trabajar para algún fin íntimo nuestro, como *objet trouvé* que incorporamos de forma útil a *nuestra* decisión de un modo u otro. Pero, para conseguirlo, debemos ser algo más que un punto matemático; tenemos que *ser alguien-,* debemos tener partes —recuerdos, planes, creencias y deseos— que habremos ido adquiriendo a lo largo del tiempo. Y será entonces cuando volverán a entrar en escena todas esas influencias causales del pasado, del exterior, dispuestas a contaminarlo todo, a ocupar el

lugar de nuestra creatividad y a usurpar el control de nuestras decisiones. Un grave dilema.

El problema, como se recordará, ya fue claramente reconocido por William James cuando preguntó: «Si un acto "libre" ha de ser una novedad absoluta que no proceda *de* mí, de mi yo previo, sino *ex nihilo,* y simplemente se asocie a mí, ¿cómo puedo yo, el yo previo, ser responsable de él?». Kane da algunos pasos importantes para responder a esta pregunta retórica con su idea de la «racionalidad plural» (Kane, 1996, capítulo 7). No queremos que nuestros actos libres sean inmotivados, inexplicables o meros destellos aleatorios carentes de fundamento. Queremos que haya razones que los justifiquen, y queremos que sean *nuestras* razones, *y* (si somos libertaristas) queremos que cumplan con la condición PA para ser libres en el sentido de que «en el tiempo *t*» «podríamos haber hecho otra cosa». Una forma de conseguirlo es tomarse la molestia de elaborar personalmente dos (o más) listas de razones *contrapuestas.* Cada cual deberá encargarse de componerlas, desarrollarlas, revisarlas, limarlas y pulirlas *localmente*, por sí mismo y sin la ayuda de nadie. Aunque tal vez hayamos podido tomar prestadas algunas ideas y detalles del exterior, los hemos hecho nuestros, de modo que no hay duda de que son listas del tipo «hágalo usted mismo». Es más, cada uno suscribe, al menos tentativamente, ambas listas de razones. (Si dejáramos de suscribir una de las dos, no habría mucho problema, ¿verdad? Habríamos tomado una decisión —tal vez incluso una decisión rápida— en favor de la otra.) De modo que cuando la deliberación llegue finalmente a su fin, vayamos por el lado que vayamos, será una opción que habremos considerado muy seriamente, hasta el límite mismo de suscribirla por entero. Nuestro acto constituye algo así como un veredicto final, una declaración que nos convierte en la clase de persona que somos (uno que se Queda o uno que se Va), y *en aquel mismo momento* podríamos haber hecho otra cosa.

La idea de la racionalidad plural —del «procesamiento paralelo», según el nombre que le da Kane recientemente (Kane, 1999)— surge de una intuición que siempre hemos tenido: es justo que se nos considere responsables por el resultado de un acto que incluye un elemento de azar o indeterminado, *si aquel resultado es el que tratábamos de obtener.* El asesino que tiene la suerte de darle al Primer ministro con un disparo lejano no queda absuelto por haber dado en el blanco por pura suerte, ni aunque sea un azar genuinamente indeterminista. Al proponer un proceso contradictorio por el que se crean dos posibilidades enfrentadas (por

ejemplo, el dilema de la mujer de negocios a propósito de hacer lo correcto o subir un escalón en su carrera), Kane garantiza que cuando una de las posibilidades falla se impone la otra, *y* que la mujer es responsable en ambos casos porque ésa es una de las cosas que estaba tratando de hacer. El hecho de que tratara de conseguir dos cosas incompatibles al mismo tiempo no demuestra que en caso de que consiguiera una de ellas no *hubiera tratado* de conseguirla.

Así pues, Kane pretende que esta inserción del indeterminismo en el conflicto de las razones, en medio del cual el agente *trata* —tipo (iii), voluntad de perseverar— de encontrar la correcta, salva al resultado, sea cual sea, de ser un golpe de suerte, un mero accidente. Cualquier agente adulto se ha enfrentado a esta clase de dilemas, de carácter moral o prudencial, y se define a partir de ellos.

> Al escoger una opción u otra en estos casos, los agentes refuerzan sus caracteres morales o prudenciales o bien sus instintos egoístas o imprudentes, según sea el caso. Se están «haciendo» a sí mismos o «formando» sus voluntades de un modo u otro, de una forma que no venía determinada por el carácter, los motivos y las circunstancias previas [...]. El hecho de que sus acciones sean, pues, una respuesta a los conflictos interiores implícitos en el carácter y los motivos previos del agente hace que su carácter y motivos puedan explicar los conflictos y el porqué de las acciones, sin explicar también el resultado de estos mismos conflictos y acciones. Los motivos y el carácter previos aportan razones para hacer una u otra cosa, pero no razones decisivas que puedan explicar qué camino tomará inevitablemente el agente (Kane, 1996, pág. 127).

La idea de que alguien que se ha enfrentado a serios dilemas de razonamiento práctico, que ha luchado contra tentaciones y resuelto disyuntivas, tiene más posibilidades de ser «dueño de sí mismo», un agente moral más responsable que alguien que simplemente haya bajado flotando felizmente por el río de la vida, aceptando las cosas tal como le venían, es una idea familiar y atractiva, pero a la que pocos filósofos han prestado atención. En la mayoría de las teorías de la libertad, la presencia de puntos de inflexión en la historia del agente no desempeña ningún papel destacado y es, de hecho, una posibilidad en buena medida ignorada, probablemente porque llama la atención sobre un embarazoso caso límite: el «asno de Buridan», que supuestamente muere de hambre porque está a la misma distancia de dos pilas de comida y no encuentra ninguna razón para ir a la derecha y no a la izquierda (o viceversa). Ya en los tiempos medievales se

hablaba de esta «libertad de la indiferencia», y siempre se ha reconocido que la mejor solución para dichos *impasses* es lanzar una moneda al aire, como complemento útil de la voluntad, podría decirse, lo cual no parece sin embargo un buen modelo de libre albedrío. Cuando los teóricos nos acercamos a una perspectiva según la cual nuestras únicas elecciones libres serían aquellas en las que también podríamos haber lanzado una moneda al aire, es que debemos haber tomado un camino equivocado. Es mejor volver atrás. Y por ello el tema sigue en buena medida ignorado. Pero Kane demuestra de manera bastante convincente que la forja del carácter que *puede* (o no) resultar de una vida de elecciones graves tomadas seriamente añade una «variedad valiosa y deseable de la libertad». Sólo hay un problema con ella, sin embargo: no necesita el indeterminismo que inspiró su creación. Es más, no puede utilizar el indeterminismo de ningún modo que lo distinga del determinismo, pues el requisito del «aquí y ahora» no sólo no está bien fundamentado: tal como veremos, es probable que sea también incoherente.

CUIDADO CON LOS MAMÍFEROS PRIMORDIALES

> *La idea básica es que la* **responsabilidad última** *reside donde se encuentra la* **causa última.**
>
> ROBERT KANE, *The Significance ofFree Will*

Tal vez piense usted que es un mamífero, y que los perros, las vacas y las ballenas son también mamíferos, pero la verdad del asunto es que no hay ningún mamífero: ¡es imposible que los haya! Ahí va un argumento filosófico para demostrarlo (tomado, con algunas alteraciones, de Sanford, 1975).

(1) Todo mamífero tiene a un mamífero por madre.

(2) En caso de que hayan existido los mamíferos, sólo ha podido ser un número finito de mamíferos.

(3) Pero la existencia de un solo mamífero supone, en razón de (1), que debe haber existido un número infinito de mamíferos, lo cual contradice (2), de modo que no puede haber habido ningún mamífero. Es una contradicción en términos.

Como sabemos perfectamente que hay mamíferos, sólo nos tomamos en serio este argumento como desafío para descubrir la falacia que se esconde detrás de él. En alguna parte debe estar el fallo, y sabemos más o menos dónde debe estar: si nos remontamos lo bastante en el árbol de familia de cualquier mamífero, llegaremos finalmente a los terápsidos, una extraña y extinta especie puente entre los reptiles y los mamíferos. Se produjo una transición gradual de los reptiles puros a los mamíferos puros, y el espacio intermedio fue cubierto por numerosos intermediarios de difícil clasificación. ¿Cómo deberíamos trazar las líneas divisorias en este espectro de cambios graduales? ¿Podemos identificar a algún mamífero, el Mamífero Primordial, que no tuviera a un mamífero por madre, y que negara de este modo la premisa (1)? ¿Con qué argumentos? Sean cuales sean estos argumentos, serán indistinguibles de los argumentos que podríamos usar para apoyar el veredicto de que dicho animal *no* era un mamífero (después de todo, su madre era una terápside). ¿Qué deberíamos hacer? Deberíamos poner coto a nuestro afán de marcar líneas divisorias. No necesitamos tales líneas. Podemos vivir con el hecho escasamente misterioso y sorprendente de que, ya ves, se produjeron una serie de cambios graduales que se fueron acumulando a lo largo de millones de años y al final dieron como resultado mamíferos puros.

Los filósofos muestran preferencia por la idea de detener el temido regreso al infinito mediante la identificación de *algo* que ponga —que debe poner— fin a la regresión: el Mamífero Primordial, en este caso. Ello les lleva a menudo a doctrinas llenas de misterio, o al menos con muchos puntos oscuros, y, por supuesto, les obliga a comprometerse en la mayoría de los casos con el esencialismo. (El Mamífero Primordial sería el primer animal dentro del conjunto de los mamíferos que tuviera todas las características esenciales de los mamíferos. Si no hay ninguna esencia definible de un mamífero, tenemos un problema. Y la biología evolutiva demuestra que no hay tales esencias.)

La teoría de la libertad de Kane postula específicamente ciertas instancias especiales donde «termina la regresión», los actos autoformativos, oAA.

> **Si queremos evitar el regreso al infinito, en algún lugar de la historia vital del agente debe haber acciones para las cuales los motivos dominantes y la voluntad a partir de la cual actúa el agente no hayan estado** *determinados de antemano* **(Kane, 1996, pág. 114).**

Tal vez podríamos detenernos a preguntar cuán frecuentes deberían ser estos momentos. ¿Una media de uno al día, uno al año o uno por década? ¿Tienden a comenzar con el nacimiento, a los 5 años o en la pubertad? Estas AA se parecen sospechosamente a los Mamíferos Primordiales. Resulta preocupante que aunque sean eventos clave en la vida de un agente moral —los ritos naturales de paso, podría decirse, a la responsabilidad adulta— resultan prácticamente imposibles de descubrir. No hay forma de distinguir una AA genuina de una pseudoAA, un arrebato de razonamiento impostor que nunca hizo uso *realmente* de la indeterminación cuántica, sino que simplemente ofreció un resultado pseudoaleatorio y, por lo tanto, determinista. Se sentirían igual desde dentro y tendrían el mismo aspecto desde fuera, por muy sofisticado que fuera nuestro equipo de observación. Tal como me sugirió Paul Oppenheim, resulta útil comparar las AA de Kane con los *eventos de especiación* que se dan en el curso de la evolución, y que sólo pueden identificarse retrospectivamente. Cada nacimiento dentro de cada cepa es un evento de especiación en potencia, puesto que cada descendiente muestra cuando menos diferencias minúsculas que lo hacen único, y cualquier diferencia podría ser el comienzo de algo que finalmente diera lugar a la especiación. El tiempo lo dirá. No hay nada *especial* en un nacimiento que se revelará más adelante como un evento de especiación.[7] De modo parecido, deberíamos desconfiar de la idea de que haya un evento —una AA— que tenga algún rasgo especial, intrínseco, local, que lo distinga de los más cercanos y explique su capacidad de fundar algo importante. ¿Acaso resulta plausible que un agente que no haya experimentado todavía ninguno de esos momentos tan especiales (sino sólo momentos parecidos, pseudoAA) no sea responsable de ninguno de sus actos? «Sí, esas cosas peludas y de sangre caliente tienen ciertamente aspecto de mamíferos, y desprenden el mismo olor y hacen

7. Algunos creacionistas contemporáneos han concedido que todos los seres vivos están relacionados por un árbol genealógico de la vida que se remonta a miles de millones de años atrás, y reconocen también que todas las transformaciones de las generaciones sucesivas dentro de una especie son el resultado de una selección natural darwiniana inconsciente, pero mantienen la esperanza de que los puntos de ramificación, las especiaciones, precisen para su explicación, si no un milagro, sí una ayuda especial por parte de algún diseñador inteligente (o el Diseñador Inteligente, pues pretenden ser neutrales en cuanto a la identidad de dicho diseñador). Esta condensación de todo lo especial en un momento mágico —o un lugar donde todo entra en conjunción— es una idea irresistible para algunos pensadores. El ejemplo más claro de ello es Michael Behe (1996); para una discusión de las falacias implícitas en dicha idea, véase Dennett (1997c).

los mismos sonidos que los mamíferos, y es posible la fertilización cruzada con otros mamíferos, pero les falta la esencia secreta; no son mamíferos, ni mucho menos.»

Consideremos a este respecto el caso de Luther. Kane dice: «Si Luther ha de ser el responsable último de su acto presente, al menos algunas de sus acciones o elecciones previas deben haber sido tales que podría haber hecho otra cosa respecto a ellas. En caso contrario, nada de lo que hubiera hecho podría haber marcado diferencia alguna en su manera de ser» (Kane, 1996, pág. 40). Así pues, tiene sentido —al menos eso pensaría uno— echar una mirada atenta a la biografía de Luther, para ver qué clase de educación tuvo, qué poderosas influencias le tenían atado, qué desgracias sufrió, y cosas por el estilo. Pero, en realidad, nada de lo que podamos descubrir acerca de tales detalles macroscópicos puede lanzar *la menor luz* sobre la pregunta de si Luther tuvo o no ninguna AA genuina durante este período. Sin duda podríamos descubrir que en diversas ocasiones hubo episodios de conflicto e introspección, y tal vez incluso podríamos confirmar que dichas «ocasiones» establecieron procesos contradictorios «caóticos» en las redes neurales de donde surgieron finalmente ciertas decisiones. Lo que no podríamos descubrir, sin embargo, es si esos conflictos disfrutaron de fuentes de variabilidad genuinamente aleatorias, y no meramente pseudoaleatorias. El precio que los libertaristas deben pagar para secuestrar sus momentos cruciales en transacciones subatómicas producidas en algún lugar privilegiado del cerebro (en el tiempo i) es convertir estas encrucijadas cruciales en indetectables tanto para el biógrafo cotidiano como para el neurocientífico cognitivista plenamente equipado. Alguien podría pensar que la diferencia entre Lutheri, que estuvo cinco años en prisión durante su adolescencia y sufrió un lavado de cerebro, y Luther_2, que tuvo una adolescencia más o menos normal con su cuota de triunfos y problemas ante el mundo, es relevante para la cuestión de si hubo AA previas a la decisión tomada por Lutherh_{oy}. Pero estas acusadas diferencias de contexto, que intuitivamente sí guardan relación con nuestra evaluación de la capacidad de Luther para la elección moral, no son en ninguna medida síntomas de la presencia o ausencia de AA. (Son tan irrelevantes para la cuestión de si hubo o no AA en Luther como lo serían los diez golpes cortos de demostración de Austin para la cuestión de si estaba o no determinado para fallar el tiro en el tiempo *t.)* Y cuando sacamos nuestros supermicroscopios y examinamos la actividad subatómica en las neuronas, cualquier cosa que veamos será igualmente irrelevante respecto a la existencia de AA.

Pero ¿acaso esta inescrutabilidad de la responsabilidad última no es un problema al que se enfrenta cualquier teoría? Como ha dicho Kane,

> **Si un joven asesino es llevado a juicio y descubrimos una vida pasada de abuso infantil y acoso de sus compañeros, debemos elaborar algún tipo de juicio respecto a qué parte de su mal carácter actual, del que su acto fue el resultado, tiene su origen en él mismo y qué parte en influencias exteriores sobre las que no tenía ningún control. Tales cuestiones son relevantes para cualquier teoría a la hora de determinar la culpabilidad o la inocencia y hasta qué punto debería mitigarse el castigo. Son cuestiones extraordinariamente difíciles de responder, con independencia de qué perspectiva se adopte sobre el libre albedrío (Kane, correspondencia personal).**

Hasta aquí, todo correcto. Las variaciones en la historia vital son sin duda relevantes para establecer variaciones en los grados de responsabilidad, tal como dice Kane, y también son difíciles de investigar, desde cualquier teoría. Pero la perspectiva libertarista de Kane requiere una investigación adicional que resulta difícil —en mi opinión, imposible— de justificar. Consideremos la cuestión desde un punto de vista estadístico: clasifiquemos a cien asesinos en función de su trasfondo social, desde los más necesitados hasta los más afortunados, para ver cuáles deberían ver mitigada su pena o recibir una plena exculpación (más adelante trataremos estas cuestiones de estrategia política). Supongamos que obtenemos los siguientes resultados: hay un 60 % que es evidente que han pasado por grandes privaciones del tipo relevante y que, por lo tanto, son candidatos no problemáticos a una mitigación sustancial de la pena; el 10 % son «casos límite» —han sufrido bastantes privaciones, pero *¿a* partir de cuánto se puede considerar que es demasiado?— y el 30 % restante da muestras de haber disfrutado de infancias normales o ejemplares, no tiene ningún signo de daños en el cerebro, etc. (véase la figura 4.9). Dichos individuos afortunados resultan ser, tras un proceso de eliminación, prácticamente indistinguibles unos de otros en todos los caracteres macroscópicos que consideramos condiciones necesarias para la responsabilidad (los caracteres que no están presentes en el otro 60 %). *Aparentemente* son todos adultos responsables. Aparentemente se encuentran entre los mejor tratados por la sociedad: los hemos criado bien, hemos cubierto sus necesidades, les hemos dado igualdad de oportunidades, etc.

La naturaleza no acostumbra a ofrecer fronteras claras, pero a veces nosotros debemos marcar una línea de estrategia política, simplemente

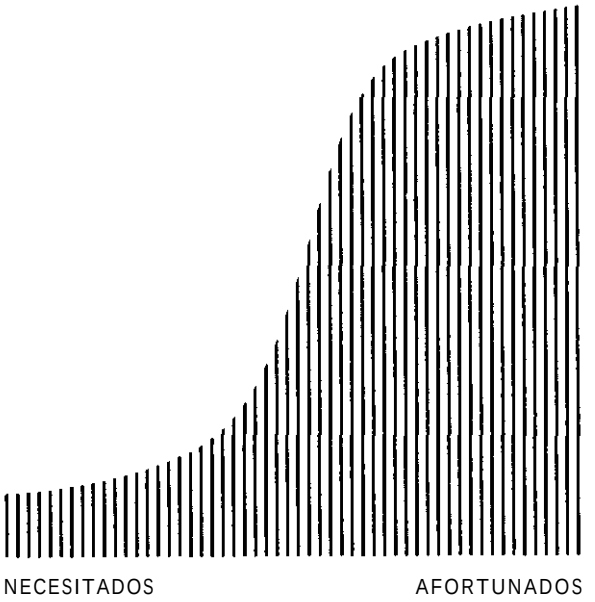

FIGURA 4.9. La distribución de los asesinos.

porque debemos tener alguna forma práctica y en principio equitativa de tratar casos concretos: en la mayoría de los Estados norteamericanos no se puede conducir hasta los 16, y no se puede beber hasta los 21, con independencia de cuán maduro sea uno para su edad. Frente al espectro de casos ilustrado por la figura 4.9, deberíamos encontrar alguna manera parcialmente arbitraria de marcar una línea en el 10 % de casos intermedios, y sin duda habría serias diferencias de opinión acerca de qué factores deberían pesar más y qué factores deberían ser ignorados. (Si la curva fuera mucho más empinada, podríamos estar agradecidos por encontrar una articulación por la que cortar la naturaleza; si fuera más gradual, nuestra tarea política sería aún más ardua.) Pero la tesis de Kane no sólo exige que reservemos nuestro juicio respecto al 10 % con derechos marginales a la mitigación de la pena, sino también respecto al 30 % de candidatos ejemplares. Un número desconocido de ellos —tal vez todos— podría resultar totalmente irresponsable, porque todas las aparentes AA de sus historias vitales fueran en realidad pseudoAA. Después de todo, Kane sostiene que ningún robot equipado únicamente con un generador pseudoaleatorio de números en su sistema puede tener ***la más mínima*** responsabilidad sobre sus actos, por más que dicho robot pudiera superar sin problema todas las pruebas macroscópicas para ser considerado humano.

(Tal robot, a diferencia de una Esposa Stepford,[8] no traicionaría su estatus de robot con una ciega obsesión por una determinada forma de actuar, puesto que los dispositivos pseudoaleatorios integrados en su facultad de razonamiento práctico conservarían eternamente su mentalidad abierta.) Según Kane, cabe perfectamente dentro de lo posible que se pudiera responsabilizar legítimamente a algunos integrantes del grupo marginal del 10 %, a pesar de sus privaciones, por haber experimentado en el pasado un modesto número de AA genuinas, mientras que algunos integrantes del grupo privilegiado del 30 % no eran candidatos válidos para la responsabilidad moral.

Tratemos de imaginar al primer acusado (el hijo de un multimillonario, puesto que va a necesitar un costoso equipo de abogados y científicos) que trata de aportar al tribunal antes de la sentencia pruebas que «demuestren» que su cerebro carecía de las indeterminaciones cuánticas necesarias para ser responsable de sus acciones, aunque haya gozado de una educación ejemplar, tenga una inteligencia superior a la media, etc. El caso pinta mal. ¿Por qué habría de pesar más el dato *metafísico* de la Responsabilidad Ultima (suponiendo que Kane haya definido una posibilidad coherente) que otros caracteres macroscópicos que pueden definirse con independencia de la cuestión del indeterminismo cuántico y que están bien fundamentados en términos de la competencia de los agentes para tomar decisiones? Es más, ¿por qué habríamos de dar valor alguno a la Responsabilidad Ultima metafísica? Si no puede servir como base para establecer una diferencia de trato entre las personas, ¿por qué habría de considerar nadie que es un tipo de libertad que merece la pena? Tal como lo expresa el propio Kane: «En resumen, descrito desde una perspectiva meramente física, *la libertad se parece mucho al azar*» (Kane, 1996, pág. 147). Y el azar tiene siempre exactamente el mismo aspecto, sea genuinamente indeterminista o meramente pseudoaleatorio o caótico.

El libertarista, igual que el esencialista en biología, está cautivado por las fronteras, en particular por las fronteras que señalan el «aquí» y el «ahora». Pero esas fronteras, al ser parcialmente interdefinibles, son siempre porosas. Supongamos que murieran las neuronas indeterministas de nuestra

8. La película de ciencia ficción de 1975 *The Stepford Wives,* de Bryan Forbes (basada en una novela de Ira Levin), presentaba una ciudad donde las esposas reales eran gradualmente sustituidas por duplicados robóticos descerebrados que dedicaban todas sus energías a la limpieza de la casa y al cuidado de sus maridos.

facultad de razonamiento práctico, lo que nos incapacitaría para cualquier futura AA. Pero supongamos también que, por fortuna para nosotros, se puede sustituir la parte dañada de nuestro cerebro por un dispositivo indeterminista implantado exactamente en el lugar adecuado dentro de la parte sana de nuestro cerebro. Una buena forma de introducir genuino indeterminismo cuántico en un dispositivo físico es usar un poco de radio en proceso de desintegración y un contador Geiger, pero tal vez no fuera sano tener implantado en el cerebro un dispositivo de radio de este tipo, de modo que éste podría quedarse en el laboratorio, rodeado de un escudo de plomo, y el cerebro podría tener acceso cuando quisiera a sus resultados por enlace radiofónico (como en mi relato «Where am I?», de *Brainstorms,* 1978). La colocación del dispositivo en el laboratorio, evidentemente, no supondría diferencia alguna, pues se hallaría *funcionalmente* en el interior del sistema; desempeñaría exactamente el mismo papel que antes realizaban las neuronas dañadas, con independencia de su localización geográfica. Pero tal vez hubiera una forma más barata y segura de conseguir el mismo efecto: podíamos usar fluctuaciones genuinamente aleatorias en la luz procedente del espacio profundo, captada por un transceptor implantado en nuestro cerebro. Como esta señal llega a la velocidad de luz, no hay forma de predecir cuáles serán las próximas fluctuaciones, aunque su fuente sea una estrella situada a años luz de distancia. Pero si no hay problema para obtener la indeterminación de una estrella lejana, ¿por qué insistir en que sea *ahora? Grabemos* una sucesión de fluctuaciones aleatorias en un dispositivo aleatorio de radio a lo largo de un siglo, e instalemos esta grabación del pasado en alguna parte de nuestro cerebro como nuestro generador de números pseudoaleatorios, para consultarlo cuando proceda.

En *Elbow Room* señalé lo irrelevante que era la diferencia entre una lotería en la que el boleto ganador fuera escogido (al azar) *después* de que todos los boletos estuvieran vendidos, y una lotería en la que el boleto ganador fuera escogido *antes* de que éstos fueran vendidos. Ambas loterías son equitativas; ambas dan a todos los participantes las mismas posibilidades de ganar:

> **Si nuestro mundo es determinista, entonces lo que hay en nosotros son generadores de números pseudoaleatorios, no dispositivos basados en un contador Geiger. O, lo que es lo mismo, si nuestro mundo es determinista, todos nuestros boletos de lotería fueron sacados de una vez, hace eones, puestos en un sobre para nosotros, y enviados a medida que los necesitábamos en el curso de nuestra vida (Dennett, 1984, pág. 121).**

Kane me ha hecho notar (correspondencia personal) que «el mecanismo generador de indeterminación debe ser sensible a la dinámica que tiene lugar en la voluntad del agente y no imponerse a ella, pues entonces estaría tomando las decisiones el mecanismo y no el agente». Su preocupación es que una fuente remota de aleatoriedad pudiera amenazar nuestra autonomía y tomar el control de nuestros procesos mentales. ¿No sería mucho más seguro —y por lo tanto más responsable— mantener el dispositivo en nuestro interior, en cierto sentido bajo nuestro ojo vigilante? No. La aleatoriedad no es más que aleatoriedad; no es una aleatoriedad *invasiva*. Los programadores introducen habitualmente remisiones al generador de números aleatorios en sus programas, sin preocuparse de que pueda escapar a su control e introducir el caos donde no se quería. Supongamos que visualizamos la dinámica del cerebro en nuestro ejemplo del Irse/Quedarse como una cresta en un paisaje de decisión, un lugar del que el explorador de la decisión deberá bajar por la pendiente norte hacia el valle del Irse o bien por la pendiente sur hacia el valle del Quedarse (véase la figura 4.10).

El paisaje está generosamente sembrado de pieles de plátano: llamadas al generador de números aleatorios, que se activan cada vez que el explorador de la decisión pasa por encima de ellas. Esto mantiene en movimiento al explorador, aleatoriamente si hace falta, lo que impide que se repita el caso del asno de Buridan, de modo que el explorador no llega nunca a quedarse encallado en la cresta y a morir en la indecisión. Las res-

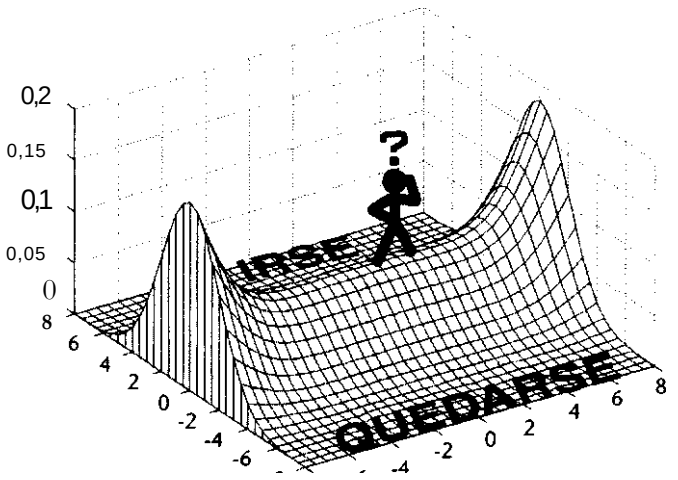

FIGURA 4.10. Cresta en un paisaje de decisión.

baladizas pieles de plátano son inofensivas, sin embargo, porque una vez que la decisión se inclina hacia uno u otro valle, encontrar una piel innecesaria sólo puede o bien causar un breve retroceso en el descenso y retrasar por un micromomento el descenso que ya se ha iniciado, o bien acelerar su caída pendiente abajo, sin poder dar marcha atrás. O, para usar otra imagen popular entre los creadores de modelos, el generador de números aleatorios simplemente «agita» o «revuelve» el paisaje sin descanso, de modo que nada puede quedarse quieto en la cresta para siempre... pero la forma del paisaje no cambia nunca, de modo que nada malo puede «invadirlo».

¿CÓMO PUEDE «DEPENDER DE MÍ»?

Un argumento popular que tiene muchas variaciones pretende demostrar la incompatibilidad entre el determinismo y la libertad (moralmente relevante) del siguiente modo:

(1) Si el determinismo es verdadero, la cuestión de si escojo Irme o Quedarme está completamente fijada por las leyes de la naturaleza y ciertos hechos del pasado remoto.

(2) No depende de mí cuáles son las leyes de la naturaleza, o qué ocurrió en el pasado remoto.

(3) Por lo tanto, la cuestión de si escojo Irme o Quedarme está completamente fijada por circunstancias que no dependen de mí.

(4) Si una acción mía no depende de mí, no es libre (en el sentido moralmente relevante).

(5) Por lo tanto, mi acción de Irme o Quedarme no es libre.

La respuesta libertarista de Kane a este convincente argumento es tratar de aislar el indeterminismo del modelo libertarista en unos pocos episodios cruciales de apertura de posibilidades «en el tiempo *t*», y espera poder localizar dichos episodios en el interior del agente, tanto a nivel espacial como temporal, para que las elecciones del agente «dependan» del agente. Pero una vez ha concedido Kane que los efectos moralmente

relevantes de estos episodios pueden verse ampliamente repartidos en el tiempo (como en el caso de Luther), ¿para qué mantener el límite del contenedor? Si cierto hecho de la infancia de Luther puede desempeñar un papel crucial en la responsabilidad del Luther adulto por su trascendental decisión de no retractarse, ¿por qué no podría tratarse de un hecho relativo a la vida de la madre, cuando Martin no era más que un embrión? Pues porque, presumiblemente, dichos hechos no tuvieron lugar dentro de Luther, sino fuera de él, en el entorno exterior a él, por más firmemente que se le impusieran, y, por lo tanto, no «dependían de» Luther. De acuerdo, pero si «el niño es padre del hombre», ¿no es el joven Luther igual de exterior respecto al Luther adulto? ¿Por qué no considerar que las disposiciones juveniles de Luther, e incluso sus ocasionales recuerdos posteriores de su infancia, no son más que influencias más bien remotas «del exterior»? Esto no es sino una versión ampliada del problema que encontramos antes en este capítulo, cuando nos preguntábamos si debíamos situar la memoria en el interior de la facultad de razonamiento práctico o dejarla fuera y que se «cargaran» partes de ella cuando la ocasión lo reclamara. Las líneas que marcamos no cumplen ninguna función clara para nosotros. Y, tal como veremos más adelante, nuestra propia agencia moral depende muchas veces de que nuestros amigos nos den un poco de ayuda, sin verse por ello desmerecida en lo más mínimo! El ideal del «hágalo usted mismo», llevado a extremos absolutistas, no es sino superstición. Es verdad que si uno consigue hacerse lo bastante pequeño, puede externalizarlo prácticamente todo. Tanto peor para los modelos que concentran todas las cuestiones en un único momento, en algún lugar del centro de un átomo. Si hay algún argumento en favor del libertarismo, tendrá que venir de algún rincón todavía por explorar, porque el mejor intento realizado hasta la fecha, el de Kane, termina en un punto muerto. Tras un examen detallado, vemos que su requisito de la Responsabilidad Ultima impone condiciones a la vez injustificadas e indetectables a las especificaciones de un agente libre. Uno puede pedir si quiere un coche con dos volantes y una brújula en el depósito de gasolina, pero eso no significa que merezca la pena tenerlos.

Así pues, ¿qué respuesta deberíamos dar al argumento incompatibilista? ¿Dónde está el paso en falso que nos excusa de aceptar la conclusión? Ahora estamos en condiciones de reconocer que comete el mismo error que el falaz argumento sobre la imposibilidad de la existencia de mamíferos. Los hechos del pasado *remoto* sin duda no «dependían de mí», pero mi decisión actual de Irme o Quedarme depende de mí porque sus «pa-

dres» —ciertos hechos en el pasado *reciente,* como las decisiones que he tomado recientemente— dependían de mí (porque *sus* «padres» dependían de mí). Y así sucesivamente, no hasta el infinito, pero sí lo bastante lejos como para darle dimensión suficiente a mi *yo* en el espacio y en el tiempo como para que haya un *mí mismo* del que puedan depender mis decisiones. La existencia de un yo moral no queda más cuestionada por el argumento incompatibilista de lo que pueda serlo la existencia de los mamíferos.

Antes de abandonar el tema del libertarismo, deberíamos preguntarnos, una vez más, cuál puede ser el sentido del proyecto. Una chispa indeterminista que salte en el momento en que tomamos nuestras decisiones más importantes no podría hacernos más flexibles, darnos más oportunidades, hacernos más autónomos o dueños de nosotros mismos en ningún sentido que pueda reconocerse *desde dentro o desde fuera,* así que, ¿qué interés habría de tener para nosotros? ¿Cómo podría marcar alguna diferencia? Sin duda, podría darse el caso de que la creencia en dicha chispa, como la creencia en Dios, cambiara toda nuestra manera de pensar sobre el mundo y sobre nuestra vida en él, incluso aunque nunca lleguemos a saber (en esta vida) si es verdad. Sí, el argumento en favor de la creencia en el indeterminismo en el marco de la acción tiene que ser algo de este tipo. Pero hay una diferencia importante. Incluso aunque nunca lleguemos a saber, ni a demostrar científicamente, la existencia de Dios, no cuesta mucho explicar por qué la creencia en un Ser supremo y bondadoso que nos protege a todos puede resultar reconfortante y darnos esperanza, fuerza moral y demás. La creencia en Dios no es lo mismo que, digamos, la creencia en Gios (una gran esfera de cobre que órbita alrededor de una estrella situada fuera de nuestro cono de luz y que lleva prominentemente estampadas en su superficie las letras Gios). Cualquiera tiene derecho a creer en Gios si eso le hace sentir mejor, pero ¿por qué habría de hacerlo? Mi cargo contra los libertaristas es que han convertido unas aspiraciones perfectamente razonables hacia unas versiones valiosas y deseables de la libertad en un anhelo por un tipo de libertad que no sería más valioso ni deseable que la comunión con Gios. Pero también es cierto que por más desviado que esté dicho anhelo, tal vez no sea inteligente tocarlo demasiado. Tal vez sea mejor que vayamos de puntillas y evitemos cualquier crítica contra este anhelo irracional e inmotivado antes de disponer de un sustituto adecuado. {*¡Detengan a ese cuervo!)* Pero si se trata de eso, ya es demasiado tarde para dar marcha atrás. Sería mejor mirar lo que se puede hacer para sacar a la gervte de su engaño.

Capítulo 4

Un examen de la mejor versión del libertarismo demuestra que no es capaz de encontrar una ubicación defendible para el indeterminismo dentro de los procesos de decisión de un agente responsable. Como no puede fundamentar su condición definitoria, podemos dejar atrás el indeterminismo y considerar condiciones más realistas para la libertad, y el modo en que podrían haber evolucionado.

Capítulo 5

Hace cuatro mil años, no había libertad en nuestro planeta, porque no había vida. ¿Qué tipos de libertad han evolucionado desde el origen de la vida, y cómo pudieron las razones de la evolución —las razones de la Madre Naturaleza— evolucionar hasta convertirse en nuestras *razones?*

NOTAS SOBRE FUENTES Y LECTURAS COMPLEMENTARIAS

Llamé la atención de los filósofos sobre la importancia del caos en *Elbow Room* (Dennett, 1984). Puede encontrarse un estudio compatibilista más reciente del papel del caos en Matt Ridley, 1999, págs. 311-313. Sobre la cuestión de dónde termina la cadena, véase *Elbow Room* (pág. 76), que también incluye discusiones sobre el caos newtoniano (págs. 151-152) y sobre el embrague móvil que marca la diferencia entre la debilidad de la voluntad y el autoengaño.

El tratamiento de los juicios rápidos en la facultad de razonamiento práctico procede de mi estudio de lo que supone *captar un chiste* en *Brainchildren* (Dennett, 1998a, pág. 86): el complejo estado disposicional de creencias que determina si alguien reirá o no ante un chiste depende de cómo complete esta persona muchos detalles que se pasan por alto al contarlo. Resultaría extraño dar el nombre de *deliberación* al proceso inconsciente que dispara una carcajada involuntaria, pero es en cualquier caso un sofisticado proceso de transformación de información.

Véase «What Happens When Someone Acts?», de David Velleman (1992), en relación con el concepto de causación por el agente de Chisholm y con una posible reducción de la misma a algo más aceptable para un naturalista, un tema que retomaremos en el capítulo 8 de este libro.

Los teóricos raramente suscriben de forma explícita el Teatro Cartesiano, pero a veces se puede conseguir que algunos salgan del armario. Para una colección de ejemplos, con comentarios, véanse mis libros y mis artículos recientes sobre la conciencia. Una imagen parecida de aislamiento en beneficio de la autoría inspira, y distorsiona, el pensamiento de algunos filósofos respecto al tema de la *comprensión.* Sobre el intento de Fred Dretske de salvar una comprensión casera genuina frente a simulacros prefabricados que pudieran comprarse e instalarse a un precio módico, véase mi «Do-it-yourself Understanding», en *Brainchildren* (Dennett, 1998a). (De acuerdo con esta perspectiva, puede *parecer* que los robots comprenden, pero no es una comprensión *suya,* porque no la realizaron ellos mismos.)

Respecto a la idea de Kane del procesamiento paralelo: en un artículo titulado «On Giving Libertarians What They Say They Want» (en Dennett, 1978) hice una sugerencia muy parecida, a partir del ejemplo (págs. 294-295) de una madre que tenía que escoger entre aceptar un puesto en la Universidad de Chicago y aceptar un trabajo en Swarthmore; ambas decisiones son racionales y, aunque la elección esté indeterminada, cuando tome su decisión, sea la que sea, habrá una buena razón para ello, y será *su* razón. Pero no me tomé la idea muy en serio, excepto como una migaja que echarles a los libertaristas. Kane ha demostrado que la subestimé.

En relación con los mamíferos: en años recientes ha surgido una literatura bastante extensa dedicada a la vaguedad y a cómo tratar con ella. Recomiendo especialmente a Diana Raffman (1996); a mí me ha convencido, pero si su exposición no le convence a usted, puede usar sus referencias bibliográficas para lo demás.

El modelo de Sobremesa [*Tabletalk*] de Robert French (1995) es un diseño tremendamente satisfactorio para tratar la clase de proceso decisorio estocástico esbozado aquí: un mundo de juguete sin ninguna relevancia moral, pero muy revelador. Véase mi Prefacio a este libro, reimpreso en *Brainchildren* (Dennett, 1998a).

Kane propone una distinción entre lo que llama versiones «epicúreas» y «no epicúreas» del indeterminismo (Kane, 1996, págs. 172-174). Un mundo de indeterminismo epicúreo contiene «bifurcaciones en la historia» (basadas en los giros aleatorios de los epicúreos) mezcladas con cosas y hechos con propiedades «determinadas». En un mundo no epicúreo hay «tanto indeterminación en las propiedades físicas como posibilidad de bifurcaciones en la historia». ¿Qué diferencia supone eso? «Un mundo epicúreo en el que ocurrieran hechos indeterminados dentro de un pasado total-

mente determinado —un mundo de azar sin indeterminación— sería un mundo de mero azar, no de libertad. No habría un "período de gestación" indeterminado para los actos libres, por decirlo así; simplemente surgirían por sorpresa del pasado determinado sin ninguna preparación previa del tipo de una tensión, conflicto o debate generador de indeterminación» (pág. 173). ¿Y qué ocurre con los modelos informáticos de conflictos no lineales, caóticos y de retroalimentación recurrente? *Aparentemente* disfrutan de «períodos de gestación» preñados de (aproximaciones digitales de) indeterminación, si se quiere, pero obtienen su pseudoindeterminismo a la manera epicúrea (con generadores de números pseudoaleatorios que insertan sus resultados en las subrutinas deterministas). No se pueden tener ambas cosas: si, siguiendo a Paul Churchland, queremos aplaudir el descubrimiento de las redes no lineales, recurrentes, con toda su apertura holista no simbólica, no rígida, debemos conceder que hay suficiente con un procesamiento algorítmico epicúreo para conseguirlo, pues así es como están hechos los modelos que se encuentran actualmente en funcionamiento.

Capítulo 5

¿De dónde viene todo el diseño?

—Perdone señor; ¿podría decirme cómo llegar a la sala de conciertos?

—¡Práctica, práctica y más práctica!

La Orquesta Sinfónica de Boston (OSB) es famosa por hacerles la vida imposible a los directores invitados hasta que demuestran su valía. Un joven director, a punto de debutar en la OSB y conocedor de la reputación de dicha orquesta, decidió ensayar un atajo para ganarse su respeto. Estaba previsto que dirigiera el estreno de una pieza contemporánea terriblemente discordante y, mientras repasaba la partitura, se le ocurrió una brillante estratagema. Encontró un *crescendo* hacia el comienzo de la pieza en el que la orquesta entera tocaba con toda su furia al menos doce notas discordantes y observó que el segundo oboe, una de las voces más suaves de la orquesta, debía tocar un si. Tomó esa parte de la partitura para el segundo oboe, e insertó cuidadosamente un signo de bemol: ahora el oboe tenía instrucciones de tocar un si bemol. En el primer ensayo, llevó briosamente a la orquesta hacia el *crescendo* previsto. «¡No!», gritó de repente, al tiempo que ordenaba parar a la orquesta. Luego, con la frente arrugada y mostrando una profunda concentración, dijo: «Alguien, veamos, sí, tiene que ser... el segundo oboe. Se suponía que debía tocar un si y usted ha tocado un si bemol». «Ni mucho menos —dijo el segundo oboe—, yo toqué un si. ¡Algún idiota había escrito un si bemol!»

LOS PRIMEROS DÍAS

Consideremos el fenómeno desde el punto de vista biológico. La Orquesta Sinfónica de Boston lleva existiendo más de un siglo, durante el cual su personal no ha dejado de cambiar, sus finanzas han sufrido altiba-

jos, su repertorio ha ido creciendo y modificándose a medida que se han ido retirando piezas viejas y explorando otras nuevas. En muchos aspectos esta vieja institución es como un organismo vivo, con una personalidad marcada, una historia particular de crecimiento, salud y enfermedad, aprendizaje y olvido, viajes por todo el mundo y retornos a casa, sustitución de «células» viejas y cansadas por nuevos reclutamientos, adaptación de su comportamiento al nicho ecológico que le ha permitido prosperar. Esta perspectiva biológica resulta atractiva y útil, pero deja fuera los rasgos más importantes y asombrosos del fenómeno. Si los biólogos de otra galaxia descubrieran la Orquesta Sinfónica de Boston, lo que más debería impresionarles no son sus notables parecidos con los animales y las plantas, sino aquello que la diferencia de ellos. Un organismo está hecho de un gran equipo de células, pero ninguna célula puede sufrir ante la perspectiva de verse humillada. Ninguna célula puede aprender a tocar el oboe, o ser responsable de la elección del director invitado de este año entre una lista de jóvenes promesas. Ninguna célula puede comprender las implicaciones de la respuesta del oboísta y anticipar el efecto catastrófico que tendrá sobre la campaña del joven director para ganarse el respeto de la orquesta.

Lo notable de la Orquesta Sinfónica de Boston (y de la miríada de instituciones y prácticas humanas) es que, por un lado, puede llegar a ser una maravilla de diseño y organización, de autosuficiencia, mientras que, por otro lado, está compuesta por una multiplicidad de individuos *autónomos,* de diferentes nacionalidades, edades, géneros, temperamentos y aspiraciones. Los miembros de la orquesta son libres para ir y venir a su antojo, de modo que el consejo de dirección debe esforzarse por garantizar las condiciones laborales y pagar lo suficiente como para mantener motivados a los miembros de la orquesta. Fijémonos en la sección de los violines. Veinte individuos de talento, pero todos diferentes. Algunos son brillantes pero perezosos, mientras que otros son unos perfeccionistas obsesivos; uno está aburrido pero trabaja a conciencia, otro está embriagado por la música, mientras que otro está soñando despierto con la posibilidad de hacer el amor con la adorable celista de delante, pero todos ellos pasan sus arcos sobre las cuerdas perfectamente al unísono, una pauta impuesta con firmeza sobre un caleidoscopio de conciencias humanas distintas. Lo que hace posible esta acción concertada es un inmenso complejo de productos culturales profundamente compartidos por los músicos, la audiencia, el compositor, los conservatorios, los bancos, las autoridades municipales, los constructores de violines, las agencias de venta de entradas, y demás. Nada en el

mundo animal puede acercarse a esta complejidad. Las mentes humanas están amuebladas —y acosadas— por miles de anticipaciones, evaluaciones, proyectos, planes, esperanzas, miedos y recuerdos que son totalmente inaccesibles a las mentes de nuestros parientes más próximos, los primates superiores. Este mundo de ideales y creaciones humanas proporciona a los seres humanos individuales unas capacidades y unas tendencias que son marcadamente distintas a las de cualquier otro ser vivo del planeta.

La libertad del pájaro para volar allí donde se le antoje es sin duda un tipo de libertad, una mejora incuestionable respecto a la libertad de la medusa de flotar por donde flota, pero un triste primo lejano de la libertad humana. Compárese el canto de los pájaros con el lenguaje humano. Ambos son unos productos magníficos de la selección natural, y ninguno de los dos es milagroso, pero el lenguaje humano revoluciona la vida, abre el mundo biológico a dimensiones completamente inaccesibles para los pájaros. La libertad humana, producto en parte de la revolución del lenguaje y la cultura, es tan diferente de la libertad del pájaro como pueda serlo el lenguaje del canto de los pájaros. Pero para comprender el fenómeno más rico, es necesario comprender primero sus componentes y predecesores más modestos. Lo que debemos hacer para comprender la libertad humana es seguir la «extraña inversión del razonamiento» de Darwin y remontarnos a la época del origen de la vida, cuando no había libertad, ni inteligencia, ni elección, sino sólo una protolibertad, una protoelección, una protointeligencia. Disponemos ya de un cuadro general de lo que ocurrió: células simples dieron origen con el tiempo a células más complejas, las cuales dieron origen a su vez a organismos multicelulares, los cuales dieron origen a su vez al complejo mundo macroscópico en el que vivimos y actuamos. Ahora debemos remontarnos y examinar algunos de los detalles más relevantes de esta procesión.

Supongamos que lo que queremos es estar vivos en el planeta Tierra. ¿Qué necesitamos? Si comenzamos a nivel molecular, no sólo necesitamos ADN, sino todas las herramientas —proteínas— moleculares para llevar a cabo los muchos pasos que requiere la replicación del ADN. Necesitamos una proteína para iniciar el proceso, otra para desenrollar la hélice, otra para fijar la cadena separada de ADN, etc., relajar el superenrollamiento, segmentar y agrupar los cromosomas, y así sucesivamente. Ninguno de estos pasos es opcional; todos son necesarios. Si falta alguna de estas proteínas, adiós. Esos mismos materiales de construcción tuvieron que ser diseñados a lo largo del tiempo. El equipo completo, que compartimos con toda la vida actual del planeta, se fue reuniendo y refinando a lo largo de

varios miles de millones de años, y sustituye a equipos más sencillos para nuestros ancestros aún más sencillos. Dependemos de nuestro equipo, y ellos dependían del suyo, pero tenemos más posibilidades que ellos gracias a las mejoras introducidas en nuestro equipo, mejoras que han sido posibles gracias a formas de integración más elevadas, las cuales a su vez han sido posibles gracias a formas aún más astutas de colisionar con las demás cosas del mundo y explotar los resultados de dichas colisiones. Cuando comenzó la vida, había sólo una forma de estar vivo. Se trataba de hacer A o morir. Ahora hay opciones: hacer A o B o C o D o... morir.

Para vivir se necesita energía. ¿Procedía del sol la primera energía explotada para la vida o de fuentes termales situadas en las profundidades de la Tierra? Esta cuestión sigue abierta en la actualidad y ha dado pie a una cantidad abrumadora de hipótesis sobre el origen de la vida que compiten por encontrar una confirmación. Fuera cual fuera su comienzo, la vida —o, en cualquier caso, la mayor parte de la vida— terminó por depender de la energía solar. Para mantenerse vivo y reproducirse era preciso estar cerca del mar o flotar en su superficie, tomando el sol. Una gran innovación se produjo cuando algunos de estos bañistas mutaron y, de este modo, «descubrieron» que, en lugar de hacerlo todo ellos mismos, podían obtener mejores resultados si invadían y desmembraban a algunos de sus vecinos para usarlos como un práctico almacén de piezas de repuesto ya construidas. Las irrupciones son lo que hace la vida interesante. Invasores e invadidos inauguraron una carrera de armamentos que llevó a nuevas variedades de ambos. Pronto —al cabo de mil millones de años aproximadamente— había muchas «maneras de ganarse la vida» (por usar la expresión de Richard Dawkins), pero esas muchas no serán nunca más que una desVaneciente cadena de actualizaciones en medio del Vasto espacio de posibilidades lógicas. Casi todas las combinaciones del material de construcción son una forma de no estar vivo.

Una de las innovaciones más importantes dentro de esta carrera armamentística de diseño competitivo fue el accidente conocido *como* la revolución eucariótica, que tuvo lugar hace varios miles de millones de años. Los primeros seres vivos, las células relativamente sencillas conocidas como procariotas, tuvieron todo el planeta para ellas durante unos tres mil millones de años, hasta que una de ellas fue invadida por una vecina, y el equipo resultante de dos fue más apto que sus primos no infectados, de modo que prosperaron y se multiplicaron, y transmitieron su capacidad de trabajar en equipo a sus descendientes. Era un primer ejemplo de un cierto tipo de cooperación: la *simbiosis,* un caso en el cual se produce una colisión entre X e Y, pero en lugar de que X destruya a Y, o al revés, o incluso peor, que

ambos resulten destruidos —el resultado habitual de las colisiones en este duro mundo—, X e Y unen sus fuerzas y crean Z, una nueva cosa más grande y versátil, con más opciones de salir adelante. Por supuesto, es posible que esto ocurriera muchas veces en el mundo procariota, pero desde la primera vez que ocurrió, la vida en el planeta cambió para siempre. Estas nuevas supercélulas, las eucariotas, vivían junto a sus primas procariotas, pero eran mucho más complejas, versátiles y competentes gracias a sus polizones. Se trataba de una cooperación involuntaria, sin duda. Los equipos eucariotas no tenían la menor noticia del trabajo en equipo que estaban realizando. No tenían —ni necesitaban— ninguna noción del arbitrario origen de su ventaja competitiva. Los primeros seres vivos eucariotas no eran multicelulares, pero abrieron un espacio de diseño que hacía posibles los organismos multicelulares, puesto que tenían las suficientes piezas de recambio como para convertirse en diferentes clases de especialistas. (Todavía estamos muy lejos de los violinistas y los oboístas, y del trabajo en equipo de la OSB, pero vamos por el buen camino.)

La revolución eucariótica nos lleva a fijarnos en el hecho de que incluso en la evolución biológica, que Darwin llamó adecuadamente «descendencia con modificación», hay mucho espacio para la transmisión *horizontal* del diseño. Los anfitriones procariotas que fueron los primeros «infectados» por sus simbióticos visitantes recibieron un inmenso regalo para mejorar su competencia, el cual había sido diseñado *en otro lugar*. Es decir, no recibieron toda su competencia por línea *vertical* de sus ancestros, a través de sus padres, sus abuelos y sus demás antepasados. En otras palabras, no recibieron toda su competencia de sus genes. Lo que sí hicieron, sin embargo, fue transmitir este regalo a toda su descendencia posterior a través de sus genes, ya que los genes de los invasores terminaron por compartir el destino de los genes del núcleo de sus anfitriones, y viajaron unos al lado de los otros hacia la siguiente generación, que fue infectada al nacer, podría decirse, con su propio complemento simbiótico. Hay un rastro claro de este doble camino que sigue siendo aún muy visible hoy en día en todas las criaturas multicelulares, incluidos nosotros. Las mitocondrias, los minúsculos orgánulos que transforman energía dentro de nuestras células, son los descendientes de aquellos invasores simbióticos y tienen sus propios genomas, su propio ADN. Nuestro ADN mitocondrial, que recibimos únicamente de nuestra madre, existe en cada una de nuestras células junto con nuestro ADN nuclear (nuestro genoma). (La reproducción sexual llegó mucho más tarde; el esperma del padre no aporta nada de sus mitocondrias en el proceso de fertilización.)

La transmisión horizontal del diseño, de información que puede ser aplicada a fines útiles, es la clave de la cultura humana, y sin duda el secreto de nuestro éxito como especie. Cada uno de nosotros es el beneficiario del trabajo de diseño realizado por incontables individuos que no son ascendientes nuestros. Cada uno de nosotros no se ve obligado a «reinventar la rueda», o el cálculo, los relojes o la forma soneto. A veces se pretende, erróneamente, que esta transmisión cultural entre individuos que no guardan lazos genéticos demuestra que la cultura humana no puede interpretarse como un fenómeno evolutivo gobernado por los principios de la teoría neodarwiniana. En realidad, tal como acabamos de ver, la transmisión horizontal de elementos útiles de diseño entre individuos no emparentados está reconocida como un elemento importante en la evolución de las formas más tempranas de vida (unicelular) del que existe una lista cada vez más amplia dé ejemplos demostrados y constituye, por lo tanto, una pieza central, no un motivo de vergüenza, para la biología evolutiva contemporánea.

La revolución eucariótica no tuvo lugar de un día para otro; antes de que pudiera asentarse fue necesario que la evolución descubriera laboriosamente las soluciones a muchos problemas. Ya en el capítulo 2 encontramos los parásitos transposons, genes renegados cuyos efectos había que frenar. El proceso que resolvió dichos conflictos intragenómicos ilustra varios aspectos importantes de la teoría darwiniana: la I+D es cara, hay que «pagar» por cada diseño, y la evolución recicla permanentemente diseños anteriores (ya pagados y copiados) para nuevos fines. Los genes de una simple célula procariota pueden *expresarse* con ayuda de un equipo relativamente sencillo de lectura de genes. Es decir, no se requiere una tecnología muy desarrollada para seguir la receta de unos genes procariotas y construir un nuevo procariota. Las más sofisticadas células eucariotas, en cambio, y no digamos los organismos multicelulares compuestos de dichos bloques de construcción más complejos, necesitan un sistema de pasos intermedios, controles y equilibrios abrumadoramente elaborados para que los genes puedan activarse y desactivarse en los momentos adecuados gracias al efecto indirecto de otros productos genéticos, y así sucesivamente. Durante algún tiempo, los biólogos se han enfrentado al clásico dilema del huevo y la gallina: ¿cómo pudo evolucionar esta elaborada maquinaria de regulación de genes? La vida multicelular no podía ni siquiera iniciar su evolución antes de que estuviera lista la mayor parte de esta maquinaria, pero aparentemente no es necesaria para la forma más simple de vida de las procariotas. ¿Qué fue lo que pagó toda esta I+D? La respuesta que comienza a ganar fuerza es que fue costeada por una guerra civil de unos

mil millones de años al comienzo de la vida procariota. Era una carrera de armamentos dentro del propio genoma, en la que los genes cívicos luchaban contra los transposons (unos gorrones que se copiaban una y otra vez en el genoma sin aportar ningún beneficio al conjunto del organismo). Esto provocó la aparición de gran cantidad de medidas y contramedidas, tales como mecanismos de silenciación y mecanismos antiaislamiento. (Cada vez conocemos mejor los detalles de dichos mecanismos, así como los detalles de los mecanismos que permitieron las unificaciones simbióticas de genomas en la revolución eucariótica, y son sin duda fascinantes, pero van más allá del alcance de este libro.) Igual que en las carreras armamentísticas actuales, el resultado fue un empate muy caro, pero los frutos de toda aquella I+D estaban disponibles para abrir nuevos caminos: la maquinaria de alta tecnología necesaria para crear formas de vida multicelulares (McDonald, 1998). Parece, pues, que nosotros somos una especie de «dividendo de la paz», igual que los ordenadores, el Teflon, el GPS y tantos otros frutos tecnológicos de la carrera armamentística dirigida por el complejo militar industrial con los dólares de nuestros impuestos.

EL DILEMA DEL PRISIONERO[1]

Pero ¿cómo son estas carreras de armamentos? ¿Qué factores gobiernan o limitan los ataques y contraataques de los diferentes «bandos» en estas competiciones? Cada vez que surge en la naturaleza algo parecido a la *cooperación* es preciso encontrar una explicación. (Es posible que tenga su origen en un feliz accidente, pero no puede ser que se mantenga por un **fe** liz accidente. Eso sería demasiado bueno para ser cierto.) Aquí es donde necesitamos la perspectiva de la teoría de juegos, y de su ejemplo clásico, el dilema del prisionero. Se trata de un sencillo «juego» para dos personas que lanza algunas pistas, unas evidentes y otras sorprendentes, sobre diferentes aspectos de nuestro mundo. El escenario básico es el siguiente. Ha sido usted encarcelado junto a otra persona en espera de juicio (digamos que bajo una falsa acusación) y el fiscal les ofrece a ambos, por separado, el mismo acuerdo: si callan los dos, sin confesar ni implicar al otro, les caerá a ambos una sentencia corta (las pruebas del Estado no son demasiado concluyentes); si usted confiesa e implica al otro y éste calla, queda usted en li-

1. Partes de esta sección están tomadas, con revisiones, de *La peligrosa idea de Darwin* **(Dennett, 1995, págs. 253-254).**

bertad, y a él le cae cadena perpetua; si ambos confiesan e implican al otro, ambos obtienen sentencias intermedias. Por supuesto, si usted calla y la otra persona confiesa, él sale libre y a usted le cae cadena perpetua.

Si ambos se niegan a colaborar, desafiando al fiscal, el resultado sería mucho mejor para los dos que si ambos confesaran, de modo que ¿por qué no hacerse la promesa de callar? (En la jerga habitual del dilema del prisionero, la opción de callar se llama *cooperar* —con el otro prisionero, por supuesto, no con el fiscal—.) Podrían prometérselo, pero ambos tendrían la tentación —cedieran o no a ella— de *traicionar*, puesto que en tal caso cada uno saldría libre, dejando al otro *primo* con todo el lío. Como se trata de un juego simétrico, la otra persona estará igual de tentada, por supuesto, de hacerle quedar a usted como un tonto con su traición. ¿Está usted dispuesto a arriesgar toda una vida en la cárcel a que el otro mantendrá su promesa? Tal vez sea más seguro traicionar, ¿verdad? De ese modo evita usted en todo caso el peor de los resultados, y tal vez incluso salga libre. Por supuesto, el otro tipo también llegará a esta conclusión, si es una idea tan buena, de modo que probablemente también optará por lo seguro y le traicionará, en cuyo caso debe usted traicionarlo para evitar la peor de las calamidades —a menos que sea un santurrón dispuesto a pasar el resto de la vida en la cárcel para salvar a un traidor—, por lo que es probable que ambos terminen por traicionarse y recibir sentencias intermedias. ¡Qué bonito sería poder superar esta línea de razonamiento y cooperar!

Lo importante es la estructura lógica del juego, no su escenario concreto, que no es sino un estímulo para la imaginación. Podemos sustituir las sentencias penales por resultados positivos (oportunidades de ganar diferentes cantidades de dinero o, digamos, de descendientes) mientras las retribuciones sigan siendo simétricas y estén ordenadas de tal modo que una traición solitaria sea más beneficiosa que la cooperación mutua, la cual dé mejores resultados que la traición mutua, la cual, a su vez, dé mejores resultados que la posición del primo, que se produce cuando el otro es el único traidor. (En entornos formales se establece otra condición: la media de las retribuciones del primo y la traición mutua no debe ser mayor que la retribución de la cooperación mutua.) *Siempre que se cumpla esta estructura en el mundo, nos encontramos ante el dilema del prisionero.*

Se han emprendido estudios teóricos a partir de juegos en los campos más diversos, como la filosofía, la psicología, la economía y la biología. En la teoría de juegos evolutiva, las retribuciones se miden en términos de descendientes y el objetivo de los modelos es examinar las condiciones bajo las cuales los diseños «cooperativos» pueden mantenerse y superar en

resultados a las instancias egoístas y traidoras que habitualmente se llevan la mejor parte. ¿Por qué es la traición la estrategia ganadora por defecto? Examinemos la matriz de retribuciones de la figura 5.1. Haga lo que haga el jugador Y, si el jugador X traiciona, obtendrá mejores resultados que si coopera. Se dice que la traición *domina* como estrategia en la situación básica. Se puede derivar matemáticamente el efecto que ello tendrá sobre los descendientes del jugador X como proporción de la generación siguiente dentro de una determinada población, y demostrarlo en simulaciones que enfrentan a agentes traidores simples de diferentes tipos con agentes cooperadores simples de diferentes tipos. Ambos interactúan en función de su tipo —los traidores siempre traicionan y los cooperadores siempre cooperan— y los resultados (en términos de número de descendientes) se van contando y acumulando a lo largo de muchas generaciones. En ausencia de circunstancias especiales que lo impidan, los traidores pronto desbordan a los cooperadores, por desgracia. Esta tendencia inevitable es el viento dominante contra el que toda la evolución de la cooperación debe luchar. La más influyente de las muchas aplicaciones del pensamiento teórico sobre juegos a la teoría evolucionista es el concepto de *estrategia evolutivamente estable,* o EEE, de John Maynard Smith, una estrategia que tal vez no sea la mejor imaginable pero que resulta insuperable por cualquier otra estrategia alternativa bajo las circunstancias dadas. Un mundo sucio donde todos traicionan siempre es una EEE en la mayoría de las circunstancias imaginables, puesto que los cooperadores pioneros que van a parar en medio de una población de este tipo son traicionados hasta la muerte en poco tiempo. Existen condiciones, sin embargo, bajo las cuales se dan otros resultados más alentadores, y esas huidas del sombrío caso general son los peldaños de la escalera que lleva hasta nosotros.

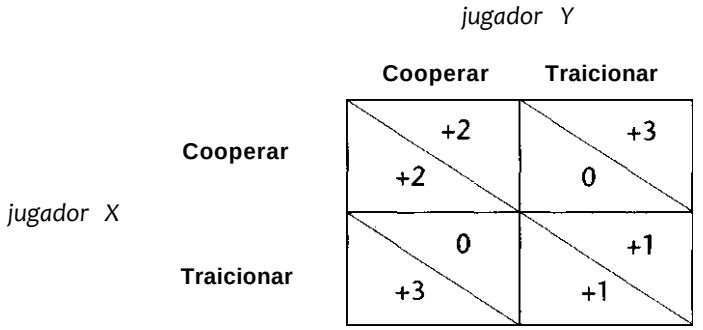

FIGURA 5.1. El dilema del prisionero.

No hay duda de que los análisis de la teoría de juegos son útiles para la teoría evolucionista. ¿Por qué, por ejemplo, son tan altos los árboles en el bosque? ¡Por *la misma razón* por la que grandes despliegues de carteles chillones compiten por nuestra atención en los centros comerciales de todo el país! Cada árbol piensa sólo en sí mismo y trata de conseguir tanto sol como sea posible. Sería mucho mejor que esas secuoyas se pusieran de acuerdo en establecer unas cuantas restricciones zonales razonables y dejaran de competir entre ellas por la luz, lo que les permitiría ahorrarse la molestia de construir esos ridículos y caros troncos, no ir más allá de ser unos matorrales bajos y ahorrativos, ¡y conseguir el mismo sol que de la otra manera! Pero no pueden ponerse de acuerdo. Bajo esas circunstancias, es inevitable que la traición produzca siempre mejores resultados que cualquier acuerdo cooperativo, de modo que si no fuera porque el suministro de luz solar es esencialmente inagotable, los árboles se verían atrapados en la «tragedia de los comunes» (Hardin, 1968). La tragedia de los comunes se produce cuando hay un recurso finito «público» o compartido que los individuos se sentirán tentados de explotar de manera egoísta más allá de la parte que les corresponde (como por ejemplo los peces comestibles). A menos que se llegue a acuerdos específicos y coercibles, el resultado tenderá a ser la destrucción del recurso. Fue el desarrollo evolutivo de controles y equilibrios coercibles lo que permitió que los genes cooperadores resistieran frente a los transposones traidores, una de las primeras innovaciones «tecnológicas» para superar el mundo aburrido y simple del egoísmo universal, la traición universal.

E PLURIBUS UNUM?[2]

El paso a la multicelularidad vino auspiciado por otra innovación en el terreno de la cooperación: resolver el problema de la solidaridad de grupo a nivel celular. Tal como señalé al comienzo del capítulo 1, cada uno de nosotros está compuesto de billones de células robóticas, cada una de las cuales dispone de un sistema completo de genes y una impresionante maquinaria interna de soporte vital. ¿Por qué se someten dichas células de manera tan altruista para el bien del equipo en conjunto? Por supuesto, lo que pasa es que han llegado a ser tremendamente dependientes las unas de

2. Esta sección contiene una versión revisada de una sección del mismo título del capítulo 16 de *La peligrosa idea de Darwin* (Dennett, 1995).

las otras, y no pueden sobrevivir mucho tiempo por sí mismas fuera del entorno particular donde acostumbran a vivir; pero ¿cómo llegaron a ser así?[5] Una de las virtudes de adoptar el «punto de vista del gen» en los estudios evolutivos es que convierte esta cuestión en un problema fundamental. La solidaridad de grupo entre las células es omnipresente en la naturaleza; al fin y al cabo, cualquier ser vivo que podamos ver a ojo descubierto está formado por células serviles y devotas. Se trata, pues, de algo «natural», pero no por ello deja de ser una proeza del diseño de proporciones superlativas, nada que los biólogos puedan dar por supuesto. Las lecciones que debemos aprender de todo ello son engañosas, sin embargo, ya que las células que nos componen pertenecen a dos categorías muy distintas.

Todas las células que componen un yo multicelular comparten un mismo ancestro; pertenecen a un único linaje celular, son hijas y nietas de un óvulo y un esperma que se unieron para formar un zigoto. Esas son las células *hospedadoras*. Las demás células, los *simbiontes*, son como las otras —son también eucariotas y procariotas—, pero cabe considerarlas forasteras porque descienden de linajes distintos. (Así pues, tenemos una simbiosis de segunda generación; la simbiosis creó nuestras células eucariotas, las cuales a su vez han hecho de anfitrionas para un sinfín de nuevos invitados.)

¿Qué es lo que conlleva la distinción entre anfitriones e invitados? La respuesta aquí, que encontrará eco en los niveles más elevados de la vida social, es que aunque el pedigrí es a menudo una buena manera de predecir la competencia futura, al final lo que cuenta es la competencia futura, con independencia del pedigrí. Por ejemplo, nuestro sistema inmunológico está compuesto por células que son ahora mismo miembros de pleno derecho del equipo anfitrión, pero iniciaron su carrera con nuestros antepasados como un ejército invasor, que fue gradualmente incorporado y con-

3. Nótese que he cedido a la costumbre de los biólogos de hablar sobre los *tipos* biológicos (o cepas o especies) como si fueran individuos. Nuestras células se han «vuelto dependientes» y, sin embargo, ninguna de mis células se ha vuelto dependiente; todas nacieron así. Las jirafas llevan eones haciendo crecer sus cuellos y los pájaros tejedores tardaron miles de años en «aprender» a construir sus nidos. El «crecer» y el «aprender» son invisibles si nos concentramos en los individuos. Tal como vimos con el surgimiento de la evitación en el capítulo 2, aunque cada individuo esté determinado para ser como es hasta el día de su muerte, el proceso en conjunto puede dar lugar a cambios, mejoras y crecimiento. Algunos filósofos recelan de esta dualidad de perspectivas —en *La peligrosa idea de Darwin* (Dennett, 1995) caractericé su escepticismo como una táctica de «dar gato por liebre»—, pero es la clave para comprender toda la I+D que se lleva a cabo a nivel evolutivo.

vertido en una tropa de guardias mercenarios, cuya identidad genética se mezcló con los linajes más antiguos a los que unieron sus fuerzas, en otro ejemplo de transmisión horizontal del diseño. La clave para comprender las pautas que siguen estas transformaciones es tratar a todas estas células robóticas como agentes individuales minúsculos, como sistemas intencionales, cada uno de los cuales posee un principio de capacidad de elección «racional». Adoptar la perspectiva intencional, pasar de la perspectiva física de los átomos componentes a la perspectiva del diseño de las simples máquinas primero, y luego a la perspectiva intencional de la agencia simple, es una táctica que da grandes resultados pero que debe utilizarse con precaución. Resulta demasiado fácil perder de vista el hecho de que hay momentos en las carreras de todos estos agentes, semiagentes y pseudosemiagentes en los que surgen —y luego pasan— oportunidades para «decidir».

Las células que me componen comparten un mismo destino, aunque algunas lo hacen en un sentido más fuerte que otras. El ADN de las células de mi dedo y mis células sanguíneas están en un callejón sin salida genético; dichas células pertenecen a la línea *somática* (el cuerpo), no a la línea *germinal* (las células sexuales). Según la memorable expresión de François Jacob, el sueño de toda célula es convertirse en dos células, pero las células de mi línea somática están condenadas a morir «sin hijos» (aparte de la posibilidad de aportar ocasionalmente un sustituto para sus vecinas muertas en acción, y de posibles avances sin precedentes en las técnicas de clonación). Como este callejón sin salida se determinó hace ya bastante tiempo, ya no hay ninguna presión, ninguna oportunidad normal, ningún «punto de elección» donde sus trayectorias intencionales —o las trayectorias de su limitada progenie— puedan alterarse. Son lo que podría llamarse un sistema *balístico* intencional, cuyos objetivos y propósitos superiores han sido fijados de una vez por todas, sin ninguna posibilidad de reorientación o variación. Son esclavos completamente sometidos al *summum bonum* del cuerpo del que forman parte. Cabe la posibilidad de que sean explotados o engañados por visitantes externos, pero bajo circunstancias normales no pueden rebelarse por sí mismos. Igual que a las Esposas Stepford, se les ha inculcado un único *summum bonum*, que no es: «Lucha por ser el número uno». Al contrario, son jugadores de equipo por naturaleza.

Cómo deben contribuir a este *summum bonum* es algo que viene implícito en su diseño, y en este sentido son fundamentalmente distintos de otras células que están «en el mismo barco»: mis visitantes simbiontes.

Los mutualistas benignos, los comensales neutrales y los parásitos nocivos que comparten el vehículo que componen entre todos —a saber, yo— tienen todos un *summum bonum* implícito en su diseño y su fin es promover sus *propias* cepas, no la mía. Por fortuna, se dan condiciones bajo las cuales puede mantenerse una *entente cordiale,* ya que, después de todo, estamos todos en el mismo barco y las condiciones bajo las cuales pueden obtener mejores resultados sin colaborar son limitadas. Pero sí tienen posibilidad de «elección». Es una cuestión abierta para ellas, a diferencia de lo que ocurre, normalmente, con las células hospedadoras.

¿Por qué? ¿Qué es lo que permite —o requiere— que las células hospedadoras sean tan serviles y al mismo tiempo den libertad a las células visitantes para rebelarse cuando surja la oportunidad de hacerlo? Por supuesto, ninguna célula es un agente racional capaz de pensar y sentir. Y ninguna posee capacidades cognitivas significativamente superiores a las demás. No es ésa la base de la teoría de juegos de la evolución. Las secuoyas tampoco son especialmente inteligentes, pero se encuentran en unas circunstancias competitivas que les obligan a traicionar, y a caer en lo que, desde *su* punto de vista (!), no es sino un desperdicio inútil. El acuerdo de cooperación mutua por el que todas podrían evitarse construir troncos elevados en el vano intento de conseguir más luz de la que les corresponde es evolutivamente incoercible.

La condición que genera la posibilidad de una elección es el «voto» inconsciente de la reproducción *diferencial.* Es la oportunidad de que se produzca una reproducción diferencial lo que permite que las cepas de nuestros visitantes «cambien de idea» o «reconsideren», al «explorar» estrategias alternativas, las opciones que tienen. Mis células hospedadoras, en cambio, han sido diseñadas de una vez para siempre por un único voto en el momento de la formación de mi zigoto. Si, gracias a una mutación, se les ocurren estrategias dominantes o egoístas, no prosperarán (en relación con sus contemporáneas), puesto que hay escasas oportunidades para la reproducción diferencial. (El cáncer puede verse como una rebelión egoísta —y destructora del vehículo— promovida por una revisión de las circunstancias normales a partir de la cual se hace posible la reproducción diferencial.)

Brian Skirms ha señalado (1994a, 1994b) un sugestivo paralelismo entre esta estrategia multicelular (otro fruto benigno de la guerra civil que creó toda la maquinaria para la lectura de genes) y la monumental *Teoría de la justicia* (1971) de John Rawls. Una cooperación normal en el marco del destino estrechamente compartido de las células de la línea somática

tiene como precondición una situación análoga a la que se da en la «posición original», el experimento mental de Rawls acerca de cómo diseñarían el Estado justo ideal unos agentes racionales si tuvieran que elegir desde detrás de lo que él llama un «velo de ignorancia». Skyrms lo llama, no sin razón, el «velo darwiniano de ignorancia». Nuestras células sexuales (esperma u óvulo) se forman por un proceso distinto de la división celular normal, o *mitosis*. Nuestras células sexuales surgen de un proceso diferente llamado *meiosis,* un proceso que construye aleatoriamente *medio* genoma (candidato a unir sus fuerzas con la otra mitad aportada por nuestro compañero o compañera) a base de elegir primero un fragmento de la «columna A» (los genes que recibimos de nuestra madre) y luego un fragmento de la «columna B» (los genes que recibimos de nuestro padre) hasta que la célula sexual tiene instalado un conjunto completo de genes —aunque sólo una copia de cada uno—, listo para probar suerte en la gran lotería del apareamiento. Pero ¿cuáles de las hijas de nuestro zigoto original están destinadas a la meiosis y cuáles a la mitosis? Esto también es una lotería.

¿Estamos hablando de una lotería *aleatoria* o *pseudoaleatoria*? Por lo que sabemos, viene a ser lo mismo que lanzar una moneda al aire, es decir, el resultado viene *determinado* por una coincidencia carente de pautas entre inescrutables influencias venidas de quién sabe dónde, y es por lo tanto predecible, en principio, por el demonio *infinito* de Laplace, pero no por las muy sensibles y variadas fuerzas selectivas que constituyen los ciegos pero efectivos tanteos del Relojero Ciego. Gracias a este mecanismo, los distintos genes paternos y maternos (en cada uno de nosotros) no pueden «conocer su destino» por adelantado en circunstancias ordinarias. La cuestión de si van a tener progenie en la línea germinal y tal vez inundar el futuro con descendientes suyos, o de si serán relegados a las aguas estériles de la esclavitud de la línea somática para el bien del cuerpo político o la corporación (nótese la etimología), es algo desconocido e imposible de conocer, de modo que los genes no tienen nada que ganar con una competición egoísta. Ese es, en todo caso, el arreglo más común. Hay ocasiones especiales en las que el Velo Darwiniano de la Ignorancia se levanta por un momento: los casos de «deriva meiótica» o de «impresión genómica» (Haig y Grafen, 1991; Haig, 1992, 2002; para una discusión véase *La peligrosa idea de Darwin* [Dennett, 1995, capítulo 9]), en los cuales las circunstancias sí permiten que tenga lugar una competición «egoísta» entre los genes —y eso es exactamente lo que ocurre, y lleva a una escalada de la carrera armamentista—. Pero en la mayoría de las circunstancias, el

«tiempo del egoísmo», en el caso de los genes, está estrictamente limitado, y en cuanto el dado —o la papeleta— ha sido lanzado, los genes están fuera de la partida hasta la próxima elección. Tal vez el primero en señalar el paralelismo fue E. G. Leigh:

> Es como si estuviéramos ante un parlamento de genes: cada uno actúa según su propio interés, pero si sus actos perjudican a los demás, éstos se combinarán para frenarlo. Las reglas de transmisión de la meiosis han evolucionado hasta convertirse en unas reglas de *fair play* cada vez más inviolables, una constitución diseñada para proteger al parlamento frente a los actos nocivos de uno solo o unos pocos. Sin embargo, en aquellos casos en los que la relación con el elemento distorsionador es tan estrecha que «seguir sus pasos» produce más beneficios que perjuicios, la selección tiende a potenciar la distorsión. Así, es necesario que una especie tenga muchos cromosomas para que, ante la aparición de un elemento distorsionador, la selección promueva su supresión en la mayor parte de los casos. Del mismo modo que un parlamento demasiado pequeño puede verse pervertido por las intrigas de unos pocos, una especie que sólo tuviera un cromosoma con vínculos muy estrechos sería una presa fácil para los elementos distorsionadores (Leigh, 1971, pág. 249).

¡Que alguien trate de describir estas pautas profundas de la naturaleza sin usar la perspectiva intencional! Las lentísimas pautas que resultan predictivas al nivel de los genes recuerdan notablemente las pautas que se revelarán predictivas en los niveles psicológico y social, y en realidad son anticipos de ellas: las oportunidades, la ignorancia y el discernimiento, la búsqueda de las mejores maniobras frente a la competencia, la evitación y la represalia, la elección y el riesgo. Hay razones que explican las medidas y contramedidas de la I+D evolutiva, aunque nada o nadie las haya valorado explícitamente. Son lo que llamo *razones virtuales* y precedieron en miles de millones de años a nuestras razones articuladas y ponderadas. Una de ellas es el principio fundamental de la evitación del daño, que rige en ambos dominios: cuando no tenemos información alguna acerca de cuál va a ser nuestro destino, la libertad de elección es perfectamente inútil.

> Y todavía hay otra forma de negarles una determinada oportunidad a las personas: mantenerlas en la ignorancia de la misma. Podríamos llamar a estas oportunidades no reconocidas ni imaginadas *oportunidades en bruto.* Si paso caminando junto a una fila de contenedores de basura, y resulta que uno de ellos contiene un bolso lleno de diamantes, me pierdo una oportunidad en bruto de hacerme rico [...]. Las oportunidades en bruto son muy abundan-

tes, pero no nos basta con ellas; cuando decimos que queremos oportunidades o posibilidades de prosperar, no queremos simplemente oportunidades en bruto. Queremos detectar nuestras oportunidades o ser informados de ellas a tiempo para que podamos actuar (Dennett, 1984, págs. 116-117).

Skyrms demuestra que cuando los elementos individuales de un grupo —sea de organismos completos o de partes de ellos— están estrechamente relacionados (son clones o casi clones) o son de otro modo aptos para el reconocimiento o el «apareamiento» por semejanza, el modelo correcto para analizar la situación deja de ser el simple dilema del prisionero, en el que domina siempre la estrategia de la traición. Es por eso por lo que nuestras células somáticas no se traicionan; son clones. Ésta es *una* de las condiciones para que los grupos —como el grupo de mis células «anfitrionas»— posean la armonía y la coordinación necesarias para comportarse de manera estable como un «organismo» o un «individuo». Pero antes de dar tres hurras y aceptarlo como nuestro modelo preferido para construir una sociedad justa, deberíamos detenernos a considerar que hay otra forma de ver a esos ciudadanos modelos, las células y los órganos de la línea somática: su modalidad específica de altruismo es la obediencia acrítica de los fanáticos, caracterizada por una lealtad al grupo fuertemente xenófoba que difícilmente podemos considerar un ideal para la emulación humana.

Nosotros, a diferencia de las células que nos componen, no nos encontramos en trayectorias balísticas; somos misiles *guiados,* capaces de alterar nuestra trayectoria en cualquier punto, abandonar objetivos, cambiar de lealtades, formarnos propósitos y luego traicionarlos, y demás. Para nosotros, cada momento es momento de decidir. Por este motivo, nos encontramos constantemente con la clase de oportunidades y dilemas sociales para los que la teoría de juegos propone un campo y unas reglas de juego, pero todavía no las soluciones. La vida es más complicada para la gente que vive en sociedad que para las células que los componen, y todavía nos queda mucha I+D que realizar —¡práctica, práctica y más práctica!— antes de llegar a la sala de conciertos.

Sin embargo, resulta alentador ver que los problemas a los que nos enfrentamos tienen precedentes que pudieron ser superados por ensayo y error. De otro modo no estaríamos aquí. El proceso de ensayo y error —incluso un ensayo y error inconsciente, capaz de conservar los progresos parciales— es poderoso. Ha sido capaz de crear novedades genuinas en el mundo; ha resuelto problemas de gran calado y superado obstáculos

que harían retroceder a cualquiera. El método de ensayo y error funciona, lo que significa que *probar* funciona: al menos una de estas pruebas tiene un historial de éxitos más que probado. Tal vez nuestros intentos no parezcan tan torpes —a pesar del determinismo— a la vista del éxito que tuvieron sus antecesores. Las propias células que nos componen son los descendientes directos de células que una vez tuvieron que resolver un gran problema de cooperación, y lo lograron.

DIGRESIÓN: LA AMENAZA DEL DETERMINISMO GENÉTICO

Después de esta siniestra disquisición sobre células y genes, combinada con otra disquisición sobre violinistas y oboístas, tal vez comience a ser hora de traer la paz a las conciencias desterrando de una vez para siempre el «espectro» del «determinismo genético». Según Stephen Jay Gould, los deterministas genéticos creen lo siguiente:

> **Si estamos programados para ser lo que somos, entonces dichos caracteres son ineluctables. Como máximo podemos canalizarlos de un modo u otro, pero no podemos cambiarlos ni por nuestra voluntad, ni por nuestra educación, ni por nuestra cultura (Gould, 1978, pág. 238).**

Si esto es el determinismo genético, entonces podemos respirar todos aliviados: no hay deterministas genéticos. Nunca he encontrado a nadie que pretenda que ni la voluntad, ni la educación, ni la cultura no puedan cambiar muchos, si no todos, de nuestros rasgos genéticamente heredados. Mi tendencia genética a la miopía se ve compensada por las gafas que llevo (pero es preciso que *quiera* llevarlas); y muchas personas de las que de otro modo padecerían una u otra enfermedad de origen genético pueden ver sus síntomas pospuestos indefinidamente con sólo recibir la debida *educación* respecto a la importancia de una determinada dieta o gracias al don *cultural* de la prescripción de una u otra medicina. Si tenemos el gen causante de la enfermedad de la fenilcetonuria, todo lo que debemos hacer para evitar sus indeseables efectos es dejar de comer alimentos que contengan fenilalanina. Tal como hemos visto, lo inevitable no depende de si reina o no el determinismo, sino de si se pueden o no tomar medidas, basadas en información que podamos obtener a tiempo, para evitar el daño previsto. Para que una decisión tenga sentido es preciso que se cumplan dos requisitos: información y posibilidad de actuar en relación con

esta información. Sin lo uno, lo otro es inútil o aún peor. En su excelente manual de genética contemporánea, Matt Ridley (1999) pone de relieve esta idea con el dramático ejemplo de la enfermedad de Huntington, que es «puro fatalismo, sin mezcla de variabilidad ambiental. Ni una buena vida, ni un buen tratamiento médico, ni una comida sana, ni una familia entregada, ni una gran fortuna pueden nada contra ella» (pág. 56). Se trata de un caso muy distinto de las muchas predisposiciones genéticas igualmente inde-seables respecto a las que sí podemos hacer algo. Y es precisamente por este motivo por lo que muchas de las personas que, a la vista de su árbol genealógico, tienen probabilidades de sufrir la enfermedad de Huntington prefieren *no* hacerse una sencilla prueba que les diría con una certeza prácticamente absoluta si la tienen o no. Pero nótese que tan pronto como se abriera una vía para tratar a aquellos que padecen la mutación de Huntington, cosa que podría suceder en el futuro, esas mismas personas serían las primeras en hacerse la prueba.

Gould y otros han declarado una firme oposición al «determinismo genético», pero dudo que nadie piense que nuestras dotaciones genéticas sean infinitamente revisables. Es prácticamente imposible que yo vaya a dar a luz, dado mi cromosoma Y. Esto es algo que no puedo cambiar ni por voluntad, ni por educación, ni por cultura (al menos no durante mi vida, pero ¿quién sabe lo que hará posible un siglo más de ciencia?). De modo que, al menos por lo que respecta al futuro previsible, algunos de mis genes fijan ciertos aspectos de mi destino sin dejarme ninguna perspectiva real de evadirme de él. Si eso es el determinismo genético, todos somos deterministas genéticos, incluido Gould. En cuanto eliminamos las caricaturas, lo que queda, en el mejor de los casos, son honestas diferencias de opinión respecto al grado de intervención que se requeriría para contrarrestar una u otra tendencia genética y, lo que es más importante, si tal intervención estaría justificada. Se trata de importantes cuestiones morales y políticas, pero muchas veces se hace casi imposible debatir sobre ellas de una forma tranquila y razonable. Un primer paso hacia la recuperación del sentido común es reconocer, como criterio práctico, que, cada vez que nos «acusen» de ser «deterministas genéticos», hay muchas probabilidades de que se trate simplemente de un caso más de *¡Detengan a ese cuervo!* y no merece la pena llevar más lejos la discusión, al menos no en esos términos. Por otro lado, ¿qué es lo que hace tan señaladamente perverso el determinismo genético? ¿No sería igual de temible el determinismo del entorno? Consideremos una definición paralela del determinismo del entorno:

Si hemos sido criados y educados en un entorno cultural determinado, entonces los rasgos que tal entorno ha impuesto sobre nosotros son inevitables. Podemos, en el mejor de los casos, canalizarlos de un modo u otro, pero no podemos cambiarlos ni por voluntad, ni por educación ulterior, ni mediante la adopción de una nueva cultura.

A menudo se cita (no sé con qué exactitud) la siguiente frase de los jesuítas: «Dadme un niño hasta los 7 años y yo os devolveré al hombre». Una exageración efectista, sin duda, pero nadie duda de que la educación temprana y otros eventos importantes de la infancia pueden tener un efecto importante sobre la vida posterior. Hay estudios, por ejemplo, que sugieren que ser rechazado por la madre en el primer año de vida o ser víctima de otros hechos traumáticos de este tipo aumentan la probabilidad de cometer un crimen violento (por ejemplo, Raine y otros, 1994). De nuevo, no debemos caer en el error de equiparar el determinismo con la inevitabilidad. Lo que debemos examinar empíricamente —y es algo que varía mucho tanto en el plano del entorno como en el genético— es si los efectos indeseados, por más grandes o profundos que puedan ser, son evitables con sólo tomar ciertas medidas. Consideremos el mal conocido como *no saber una palabra de chino.* Yo personalmente lo padezco, totalmente por causa de influencias del entorno durante mi infancia temprana (mis genes no tenían nada que ver con ello, al menos directamente). Si tuviera que mudarme a China, sin embargo, pronto estaría «curado», sólo con que pusiera algo de esfuerzo de mi parte, aunque sin duda conservaría signos profundos e insuperables de mi deficiencia, fácilmente detectables para cualquier hablante nativo de chino, durante el resto de mi vida. Pero ciertamente conseguiría arreglármelas lo bastante bien en chino como para ser responsable de las acciones que pudiera emprender bajo la influencia de los hablantes de chino que me encontrara.

¿Acaso no es cierto que todo lo que no viene determinado por nuestros genes debe venir determinado por nuestro entorno? ¿Qué más puede haber? Está la Naturaleza y está la Crianza. ¿Hay alguna otra X, algún factor ulterior que contribuya a lo que somos? Está el Azar. La Suerte. Ya hemos visto en los capítulos 3 y 4 que este ingrediente extra es importante, pero no tiene por qué venir de las entrañas cuánticas de nuestros átomos o de ninguna estrella lejana. Nos rodea por todas partes en los eventos azarosos y carentes de causa de nuestro ruidoso mundo, que llenan automáticamente todas las lagunas de especificación que dejan sin fijar nuestros genes o las causas más prominentes que operan en nuestro en-

torno. Esto es particularmente evidente en la formación de los billones de conexiones en nuestro cerebro. Hace años que se sabe que el genoma humano, a pesar de su extensión, es con mucho demasiado pequeño para *especificar* (en su receta genética) todas las conexiones que se forman entre las neuronas. Lo que ocurre es que los genes especifican procesos que disparan grandes aumentos en la población de neuronas —muchas más de las que nuestros cerebros usarán nunca—, las cuales despliegan terminaciones de manera aleatoria (pseudoaleatoria, por supuesto) hasta que en muchos casos conectan *casualmente* con otras neuronas de formas que resultan útiles en un sentido *detectable* (detectable para los inconscientes procesos de poda del cerebro). Estas conexiones ganadoras tienden a sobrevivir, mientras que las conexiones perdedoras mueren, para ser luego desmanteladas de modo que sus partes puedan reciclarse unos días más tarde en la próxima hornada de crecimientos neuronales. Este entorno selectivo en el interior del cerebro (especialmente en el cerebro del feto, mucho antes de que llegue al entorno exterior) no especifica las conexiones finales más de lo que puedan hacerlo los genes; hay factores, tanto en los genes como en el entorno, que influyen y podan su crecimiento, pero hay mucho que queda en manos del azar.

Cuando recientemente se publicó el genoma humano y se anunció que «sólo» tenemos alrededor de 30.000 genes (según los criterios actuales sobre cómo identificar y contabilizar los genes), no los 100.000 que habían estimado algunos expertos, corrió un divertido suspiro de alivio entre la prensa. ¡Uf! No somos meros productos de nuestros genes: ¡«nosotros» aportamos todas las especificaciones que de otro modo habrían «fijado» aquellos 70.000 genes! Pero cabe preguntarse: ¿cómo vamos a cumplir *nosotros* con esta tarea? ¿No nos exponemos a una amenaza igual por parte de nuestro temible entorno, de la vieja Crianza con sus insidiosas técnicas de adoctrinamiento? ¿Cuando la Naturaleza y la Crianza hayan hecho su trabajo, quedará algo para que yo pueda ser *yo?* (Si uno se hace realmente pequeño, puede externalizarlo prácticamente todo.)

¿Acaso importa cuál sea el compromiso concreto si, sea como sea, son nuestros genes y nuestro entorno (incluido el azar) los que se reparten el botín y «fijan» nuestros caracteres? Tal vez parezca que el entorno es una fuente más benigna de determinación, pues, después de todo, «podemos cambiar el entorno». Eso es cierto, pero no podemos cambiar el entorno *pasado* de una persona más de lo que podemos cambiar a sus padres, y los ajustes futuros en el entorno pueden dirigirse con igual firmeza a compensar limitaciones genéticas previas como limitaciones ambientales previas. Y

en la actualidad estamos a un paso de poder ajustar el futuro genético casi con la misma facilidad que el futuro entorno. Supongamos que sabemos que uno de nuestros hijos tendrá un problema que puede mitigarse con un reajuste en sus genes o bien en su entorno. Puede haber muchas razones válidas para preferir un tratamiento que otro, pero ciertamente no es evidente que una de esas opciones deba ser descartada sobre bases morales o metafísicas. Suponga, por proponer un caso imaginario que probablemente se verá superado pronto por la realidad, que usted es un inuit convencido para quien la vida por encima del Círculo Ártico es la única que vale la pena vivir, y supongamos que le dicen que sus hijos estarán genéticamente incapacitados para vivir en un entorno de este tipo. Puede mudarse a los trópicos, donde sus hijos estarán bien —al precio de renunciar a su herencia *ambiental*—, o puede reajustar su genoma para que puedan seguir viviendo en el mundo Ártico, al precio (si es que puede considerarse así) de la pérdida de algún aspecto de su herencia genética «natural».

La cuestión no es el determinismo, sea genético o del entorno, o de ambos a la vez; la cuestión es *qué podemos cambiar,* sea o no determinista el mundo. Jared Diamond ofrece una fascinante perspectiva sobre la equívoca cuestión del determinismo genético en su magnífico libro *Armas, gérmenes y acero* (1997). La cuestión que plantea Diamond, y que en buena medida responde, es por qué los «occidentales» (los europeos o los euroasiáticos) han logrado conquistar, colonizar y en sentido amplio dominar a la gente del «Tercer Mundo», y no ha sucedido al revés. ¿Por qué las poblaciones humanas de América o África, por ejemplo, no crearon imperios mundiales capaces de invadir, masacrar y someter a los europeos? ¿Es la respuesta... la *genética*? ¿Acaso ha demostrado la ciencia que la fuente última del dominio de Occidente reside en nuestros genes? Cuando ven planteada por primera vez esta pregunta, muchas personas —incluso científicos muy competentes— dan por supuesto que Diamond, por el mero hecho de plantearla, debe suscribir alguna hipótesis racista acerca de la superioridad genética de los europeos. Tanto les inquieta su sospecha que les cuesta mucho comprender que lo que dice Diamond (y debe esforzarse mucho para dejarlo claro) viene a ser lo contrario: la explicación secreta no reside en nuestros genes, los genes humanos, pero sí en buena medida en otros genes: los genes de las plantas y los animales que fueron los antepasados silvestres de todas las especies domesticadas que se emplean en la agricultura humana.

Los guardias de prisión tienen un dicho: si algo puede ocurrir, ocurrirá. Quieren decir con ello que si existe algún fallo en la seguridad, alguna

prohibición o vigilancia inefectivas o algún punto débil en las barreras, pronto serán encontrados y explotados hasta donde sea posible por los prisioneros. ¿Por qué? La perspectiva intencional lo deja bien claro: los prisioneros son sistemas intencionales frustrados que se caracterizan por la inteligencia y la astucia; eso significa mucho deseo informado con una gran cantidad de tiempo libre para explorar su entorno. Su procedimiento de búsqueda será tan bueno como exhaustivo, y sabrán distinguir las mejores estrategias de las que no lo son tanto. No hay duda de que encontrarán cualquier cosa que se pueda encontrar. Diamond parte del mismo criterio, y supone que la gente de cualquier lugar del mundo ha sido siempre más o menos igual de lista, eficiente y oportunista, igual de disciplinada, previsora, como la de cualquier otra parte del mundo, y, por lo tanto, que la gente ha encontrado siempre lo que había por encontrar. En una primera aproximación, puede decirse que *todas las especies salvajes domesticabas han sido domesticadas.* La razón por la que los euroasiáticos tomaron la delantera en el terreno tecnológico fue porque tomaron la delantera en el terreno agrícola, y lo consiguieron porque entre las plantas y los animales salvajes que tenían a su alrededor hace diez mil años había candidatos ideales para la domesticación. Había hierbas que eran genéticamente fáciles de convertir en superplantas: sólo hacían falta unas pocas mutaciones para que pasaran a ser nutritivos y generosos cereales, algo que podía conseguirse más o menos por accidente; también había animales que por su naturaleza social estaban genéticamente cerca de convertirse en animales de pastoreo y que se reproducían fácilmente en cautividad. (El maíz costó más de domesticar en el hemisferio occidental en parte porque debía recorrer una mayor distancia genética respecto a su precursor silvestre.) La mayor parte de los factores de selección que permitieron cubrir este camino, antes de la llegada de la agronomía moderna, correspondieron sin duda a lo que Darwin llamaba la «selección inconsciente»: las tendencias en gran medida involuntarias y ciertamente no informadas que estaban implícitas en los patrones de comportamiento de unas personas que tenían una perspectiva tremendamente limitada sobre lo que hacían y por qué lo hacían. Las causas principales, los factores que «fijaron» las oportunidades de la gente en cada lugar, fueron ante todo los accidentes biogeográficos y, por lo tanto, el entorno. Gracias a haber vivido durante milenios rodeados de muchas variedades de animales domesticados, los eurasiáticos desarrollaron además inmunidad hacia diversos agentes patógenos susceptibles de transmisión entre animales y humanos —aquí sí que hay que destacar un papel importante que corresponde a los genes huma-

nos, confirmado más allá de cualquier sombra de duda— y cuando, gracias a su tecnología estuvieron en condiciones de recorrer largas distancias y encontrarse con otros pueblos, sus gérmenes hicieron muchas veces tanto daño como sus armas y su acero.

¿Qué podemos decir de la tesis de Diamond? ¿Es un peligroso determinista genético, o un peligroso determinista ambiental? No es ninguna de las dos cosas, naturalmente, pues las dos modalidades de hombre del saco son tan míticas como los hombres-lobo. Al aumentar la información que tenemos a nuestro alcance acerca de las diversas causas que llevaron a las condiciones que limitan nuestras oportunidades actuales, ha aumentado nuestra capacidad para evitar y prevenir lo que queramos prevenir. El conocimiento del rol que desempeñan nuestros genes y los genes de las especies que nos rodean no es un enemigo de la libertad humana, sino uno de sus mejores aliados.

GRADOS DE LIBERTAD Y LA BÚSQUEDA DE LA VERDAD

Para poder percibir las «decisiones» que toman los *linajes* (sean de células parásitas o sean de secuoyas) hay que mirar de la manera adecuada. Es necesario adoptar la perspectiva intencional hacia esas curiosas agregaciones de materia, rebobinar el tiempo hacia adelante y esperar a que los patrones de alto nivel emerjan, como en efecto lo hacen, en medio de la montaña de datos con agradable predictibilidad. Las decisiones más reconocibles, realizadas en tiempo real por individuos compactos y fácilmente identificables, tuvieron que esperar al surgimiento de la locomoción. Sí, los árboles pueden «decidir» que la primavera ha llegado y que es hora de hacer brotar sus flores, y las almejas pueden «decidir» cerrarse cuando perciben un golpe alarmante sobre sus conchas, pero esas elecciones son tan rudimentarias, tan parecidas al funcionamiento de un simple interruptor, que sólo las consideramos decisiones por cortesía. Pero incluso los interruptores, que se limitan a encenderse o apagarse como consecuencia de ciertos cambios en el entorno, dan lugar a un cierto *grado de libertad,* tal como dicen los ingenieros, y requieren por lo tanto un cierto grado de control, sea del tipo que sea. Un sistema posee un cierto grado de libertad cuando hay un conjunto de posibilidades diferentes y la actualización de una u otra de estas posibilidades depende de la función o del interruptor que controla esta libertad. Los interruptores (que pueden tener sólo dos o bien múltiples posiciones) pueden estar conectados unos a otros en serie,

en paralelo o en sistemas que combinen ambos tipos de conexión. A medida que estos sistemas proliferan, y forman redes de interruptores cada vez más grandes, los grados de libertad se multiplican de forma asombrosa y los problemas de control se vuelven complejos y no lineales. Cualquier linaje equipado con una estructura de este tipo se enfrenta a un problema: ¿qué información *debería* modular el paso por esta red de caminos que se bifurcan en un espacio multidimensional de posibilidades? Para eso está el cerebro.

Un cerebro, con sus bancos de inputs sensoriales y outputs motores, es un dispositivo centralizado para rastrear el entorno pretérito en busca de información que pueda luego refinar hasta dar con las pepitas de oro de unas expectativas válidas sobre el futuro. Luego cada uno puede usar estas expectativas que tanto trabajo le ha costado conseguir para modular sus elecciones, mejor de lo que sus congéneres modulan las *suyas*. La velocidad es esencial, puesto que el entorno cambia constantemente y está plagado de competidores, pero también lo es la precisión (puesto que una de las opciones de los competidores es el camuflaje) y, por lo tanto, la eficiencia (puesto que todo tiene un precio y al final tiene que dar resultados). Estas condiciones evolutivas dan lugar a un conjunto de compromisos que premian una atención sensorial rápida, de alta precisión y muy focalizada. La carrera armamentística garantiza que cada especie ignorará tantos aspectos como pueda de su entorno, una estrategia peligrosa que puede dar lugar a sorpresas desagradables en el futuro, cuando una variable del entorno que hasta el momento había resultado trivial cobre de repente una relevancia crucial.

Esta perspectiva de orden superior sobre un entorno rico en novedades no previsibles y sin embargo relevantes plantea otra posible apuesta: ¿saldrá a cuenta que algún linaje invierta en *aprendizaje*? Dicha apuesta tiene un coste sustancial de carácter general: es preciso instalar cierta maquinaria para hacer posibles unas redes de interruptores que puedan rediseñarse en tiempo real, durante la vida del propio individuo, para que éste pueda ajustar sus funciones de control en respuesta a las nuevas pautas que detecte en el mundo. Recordemos la distinción de Drescher (1991) entre las *máquinas de situación-acción* y las *máquinas de elección* mencionada en el capítulo 2. Las máquinas de situación-acción consisten en un sistema de interruptores relativamente sencillos, cada uno de los cuales encarna lo que viene a ser una especie de regla de respuesta al entorno: *si encuentras la condición C, haz A*. Estos interruptores son eficientes para organismos relativamente simples cuyo comportamiento está especificado desde el nacimiento. Las máquinas de elección tienen un conjunto de me-

canismos distintos, que encarnan predicciones del tipo: *si encuentras la condición C, hacer A tendría el resultado Z (con la probabilidad p).* Generan varias o muchas predicciones de este tipo, y luego las evalúan (usando cualesquiera valores que tengan previamente o hayan desarrollado ellos mismos), un funcionamiento que resulta eficiente para organismos diseñados para aprender en el curso de su vida. Un organismo puede tener instaladas ambas clases de maquinaria y confiar en la primera para realizar elecciones rápidas y poco refinadas para salvar la vida, y en la última para pensar seriamente acerca del futuro: una rudimentaria facultad de razonamiento práctico.

Esta sofisticada maquinaria para el aprendizaje sólo saldrá a cuenta si hay suficientes ocasiones para aprender (y si el aprendizaje tiende a ir en la dirección de adquirir hábitos buenos, no hábitos malos, por supuesto). ¿De cuántas ocasiones estamos hablando? Eso depende de las circunstancias, pero no hay duda de que a menudo no las hay. «Úsalo o piérdelo» es un lema que tiene muchas aplicaciones en el mundo animal. Por ejemplo, los cerebros de los animales domesticados son significativamente más pequeños que los cerebros de sus parientes más próximos en estado salvaje, y eso no es sólo un producto secundario de la selección en busca de una mayor masa muscular en animales criados para buscarse su propio alimento. Los animales domesticados pueden permitirse ser estúpidos sin dejar de tener una descendencia muy numerosa, pues lo que han hecho ha sido delegar muchas de sus tareas cognitivas en otra especie, nosotros, de la que se han convertido en parásitos. Del mismo modo que las tenias han «decidido» confiar en nosotros para realizar todas sus tareas de locomoción y búsqueda de alimento, gracias a lo cual pueden simplificar drásticamente unos sistemas nerviosos que ya no necesitan, los animales domesticados se verían en dificultades si no pudieran contar con sus anfitriones humanos. No son *en-*¿/oparásitos que vivan dentro de nosotros, pero siguen siendo parásitos.

Estamos ya muy cerca de la libertad del pájaro, que puede volar e ir allí donde quiera. ¿Por qué quiere el pájaro ir allí donde quiera? Tiene sus motivos. Sus razones están encarnadas en la configuración de los interruptores de su cerebro, y vienen avaladas, a largo plazo, por su continuada supervivencia. Principalmente, las cosas sobre las que el pájaro se toma la molestia de reunir información son las más relevantes para su bienestar inmediato. Cuanta más presión de algún taimado competidor hayan sufrido sus ancestros en los últimos tiempos, más probable es que lleve instalado un costoso equipo para contrarrestar sus amenazas. Cuando los marineros llegaron por primera vez con sus barcos a las remotas islas del Pacífico, habitadas por

pájaros cuyos ancestros llevaban miles de años sin ver a ningún depredador, encontraron ejemplares tan poco curiosos, tan poco temerosos de las grandes cosas en movimiento que se les acercaban, que los marineros podían tender la mano y cogerlos. Esos pájaros eran perfectamente capaces de volar, pero no hacía falta tomar la menor precaución para capturarlos. Podían volar y desplazarse hasta donde quisieran, pero les faltaba picardía; había razones en el aire para hacer ciertas cosas, pero ellos no eran lo bastante listos como para captarlas. Tenían gran cantidad de oportunidades en bruto de salvarse, pero les faltaba la información necesaria para hacerlas efectivas. Por supuesto, la mayor parte de esas especies de pájaros están hoy extinguidas.

La carrera armamentística entre el depredador y la presa, así como la competición entre congéneres por el apareamiento y por los medios que llevan al apareamiento —comida, refugio, territorio, posición dentro del grupo, etc.—, han dado lugar en nuestra biosfera a cientos de miles de años de I+D en un amplio espectro de procesos paralelos y simultáneos protagonizados por millones de especies. En este mismo instante, billones de organismos de este planeta están jugando al escondite. Pero para ellos no es sólo un juego; es una cuestión de vida o muerte. *Hacerlo bien,* no cometer errores, es importante para ellos —en realidad no hay nada más importante—, pero en general no son conscientes de ello. Son los beneficiarios de un equipo exquisitamente diseñado para hacer bien lo que deben hacer, pero cuando este equipo no funciona como es debido y lo hace mal, no disponen de ningún recurso, por lo general, para darse cuenta de ello, y ya no digamos para lamentarlo. Siguen adelante, sin darse cuenta del problema. La diferencia entre lo que parecen las cosas y lo que son realmente es una diferencia tan crucial para ellos como pueda serlo para nosotros, pero les resulta en buena medida desconocida. El *reconocimiento* de la diferencia entre la apariencia y la realidad es un descubrimiento humano. Sólo unas pocas especies —algunos primates, algunos cetáceos, tal vez incluso algunos pájaros— muestran signos de apreciar el fenómeno de la «falsa creencia» (*hacerlo mal).* Muestran sensibilidad a los errores de otros en cuanto errores, y tal vez incluso cierta sensibilidad hacia los suyos propios, pero carecen de la capacidad de reflexión necesaria para *concentrarse* en esta posibilidad y, por lo tanto, no pueden usar esta sensibilidad para el diseño deliberado de reparaciones o mejoras en su propio equipo de búsqueda o de ocultación. Esa brecha entre la apariencia y la realidad es algo que sólo los seres humanos hemos logrado salvar.

Somos la especie que descubrió la duda. ¿Hay bastante comida almacenada para el invierno? ¿He calculado mal? ¿Me está engañando mi pa-

reja? ¿Debería haberme desplazado hacia el sur? ¿Es seguro entrar en esta cueva? Otras criaturas se muestran a veces visiblemente agitadas por sus propias incertidumbres respecto a cuestiones parecidas, pero como no pueden *hacerse a sí mismas esas preguntas,* no pueden articular sus problemas por sí mismas o dar pasos para acercarse más a la verdad. Están atrapadas en un mundo de apariencias, sacan tanto partido como pueden a lo que parecen ser las cosas y raras veces se preocupan, si es que llegan a hacerlo nunca, por si son verdaderamente tal como parecen. Sólo a nosotros pueden asaltarnos las dudas, y sólo a nosotros nos ha aguijoneado esta inquietud epistémica para buscar formas de remediarla: mejores métodos para la búsqueda de la verdad. En nuestro intento de mejorar el control que tenemos sobre nuestras fuentes de alimento, nuestros territorios, nuestras familias, nuestros enemigos, descubrimos los beneficios de hablar con los demás, hacer preguntas, transmitir la tradición. Inventamos la cultura.

Es la cultura lo que nos da el punto de apoyo que nos permite elevarnos a un territorio nuevo. La cultura nos da una atalaya desde la cual podemos ver cómo cambiar las trayectorias hacia el futuro que han diseñado las investigaciones ciegas de nuestros genes. Tal como ha dicho Richard Dawkins: «Lo importante es que no existe ninguna razón general para suponer que las influencias genéticas sean más irreversibles que las del entorno» (Dawkins, 1982, pág. 13). Pero para revertir esta clase de influencias, debemos ser capaces de reconocerlas y comprenderlas. Sólo nosotros, los seres humanos, disponemos del conocimiento a largo plazo necesario para identificar y evitar las trampas que se esconden en los caminos proyectados por nuestros genes carentes de previsión. El conocimiento compartido es la clave para aumentar nuestra libertad respecto al «determinismo genético».

No hemos llegado aún a la sala de conciertos, pero estamos cada vez más cerca.

Capítulo 5

La sabiduría implícita en el diseño de las formas de vida multicelulares sólo se entiende adecuadamente cuando se adopta la perspectiva intencional respecto al conjunto del proceso evolutivo. Desde esta perspectiva podemos discernir las razones virtuales que hay detrás de las «elecciones» cooperativas en juegos de no suma cero que han guiado los procesos evolutivos de

I+D hacia la creación de agentes racionales cada vez más sofisticados, y hacia la multiplicación de la capacidad de las formas de vida para reconocer y reaccionar ante las oportunidades. Cuando nos olvidamos del espantajo del «determinismo genético», vemos como la evolución guiada por la selección natural hace posibles unos niveles cada vez mayores de libertad, aunque no se trata aún de la libertad de la agencia humana.

Capítulo 6

La cultura humana no es ni un milagro ni una más de las herramientas que nos proporcionan nuestros genes para mejorar su propia competencia. Para comprender cómo es posible que una persona sea a la vez creadora de cultura y una creación de la misma debemos examinar el proceso evolutivo que ha llevado a lo largo de múltiples etapas al surgimiento de la cultura y la sociabilidad humana.

NOTAS SOBRE FUENTES Y LECTURAS COMPLEMENTARIAS

Pueden encontrarse desarrollos más extensos de las ideas presentadas en este capítulo en *La peligrosa idea de Darivin* (Dennett, 1995), de donde proceden algunos párrafos del capítulo. *Games, Sex and Evolution,* de John Maynard Smith (1988; sobre todo los capítulos 21 y 22) es una excelente introducción a la aplicación de la teoría de juegos a la evolución, igual que la edición revisada de *El gen egoísta* (1976), de Richard Dawkins. *Evolution of the Social Contract* (1996), de Brian Skyrms, pasa revista a investigaciones más recientes. Para un atractivo repaso general a la perspectiva esbozada en este capítulo, véase *Nonzero: The Logic of Human Destiny* (2000), de Robert Wright.

Nuestra comprensión de los procesos evolutivos aquí descritos, en especial de los conflictos entre genes que pueden describirse desde una perspectiva intencional, aumenta con gran rapidez. Muchas de las tesis específicas que se sostienen hoy (como el número de genes del genoma humano) puede que queden canceladas mañana, pero el esqueleto de la teoría y la evidencia que está en la base de la biología evolutiva es notablemente firme y resistente. Un libro excelente, aunque difícil, que pasa revista a los pasos en la transición desde las formas de vida más sencillas hasta las sociedades humanas es *The Major Transitions in Evolution* (1995), de

Maynard Smith y Eórs Szathmáry; una versión más sencilla es su libro de 1999, *Ocho hitos de la evolución: del origen de la vida a la aparición del lenguaje*. Para una revisión autorizada del estado de la cuestión hacia finales de 2000, véase *Evolution: From Molecules to Ecosystems,* editado por Andrés Moya y Enrique Font (de próxima aparición), donde se recogen una serie de estudios sobre temas tales como la evolución de la multicelularidad, los conflictos que pueden surgir a pesar del destino en buena medida compartido de los genes mitocondriales y nucleares, de los compromisos coste-beneficio de la simbiosis, y muchos otros temas fascinantes.

La distinción de Drescher entre las máquinas de situación-acción y las máquinas de elección clarifica de forma útil la distinción (y en parte coincide con ella) que propongo entre criaturas skinnerianas y popperianas (Dennett, 1975, 1995, 1996a).

Capítulo 6

La evolución de mentes abiertas

Los seres humanos no son simplemente brutos inteligentes, agentes ingeniosos que miran por sí mismos en un mundo peligroso, y tampoco son un mero rebaño de animales que se apiñan en busca de un beneficio común que no necesitan comprender. Nuestra sociabilidad es un fenómeno complejo, preñado de desarrollos de gran alcance como el reconocimiento mutuo (del reconocimiento del reconocimiento...), y que abre, por lo tanto, toda clase de oportunidades para actividades específicamente humanas, como el establecimiento de promesas y su ruptura, la veneración y la calumnia, el castigo y el honor, el engaño y el autoengaño. Es esta complejidad de nuestro entorno la que empuja a nuestros sistemas de control, nuestras mentes, a desarrollar su propia complejidad para que podamos hacer frente de forma efectiva al mundo que nos rodea (si somos normales). Hay seres humanos que tienen la desgracia de no poder hacerlo, por una razón u otra, y deben vivir entre nosotros con un estatus reducido, como si fueran mascotas, en el mejor de los casos: cuidadas y respetadas, controladas si es necesario, amadas y amantes a su manera limitada, pero que no participan plenamente en el mundo social y, por supuesto, carecen de la libertad moralmente relevante. Los límites problemáticos que se dan entre esas personas y el resto de nosotros, y las cuestiones extraordinariamente difíciles que se plantean cuando los individuos están a un paso de la promoción o la degradación serán el tema de un capítulo posterior, pero para preparar el terreno debemos considerar antes con más detalle cómo evolucionó esta complejidad única de la sociedad y la psique humanas.

CÓMO LOS SIMBIONTES CULTURALES CONVIRTIERON A LOS PRIMATES EN PERSONAS

> *Una araña realiza operaciones parecidas a las de un tejedor, y una abeja deja en ridículo a muchos arquitectos al construir sus celdas. Pero lo que distingue al peor arquitecto de la mejor de las abejas es lo siguiente: que el arquitecto levanta su estructura con la imaginación antes de poner sus cimientos en la realidad.*
>
> KARL MARX, *El capital*

> *La cultura hace las cosas más fáciles, o directamente posibles. Y algunos de los cambios que ha traído parecen más inexorables («evolutivos») que otros.*
>
> JOHN MAYNARD SMITH,
> **«Models of Cultural and Genetic Change»**

Entre las especies que ponen los huevos y se van, sin compartir nunca el mismo medio con su descendencia, los genes son prácticamente la única vía de transmisión vertical de rasgos hereditarios. Prácticamente, pero no del todo, como puede verse en el siguiente ejemplo: tomemos una especie de mariposa que normalmente pone sus huevos en las hojas de una especie concreta de planta, y consideremos lo que puede ocurrir cuando una hembra pone accidentalmente sus huevos en la hoja de una planta de otra especie. Es probable que el (principal) gen responsable de este hábito en la puesta de los huevos funcione a base de hacer que las larvas «graben» la primera hoja que observen al salir del huevo. La descendencia de esta mariposa anormal repetirá su mismo «error» e instintivamente pondrá *sus* huevos en hojas que se parezcan a la hoja donde nació. Si su error resulta ser un error feliz, es posible que su línea prospere mientras otras perecen: la preferencia por la nueva hoja será una adaptación que no comportará *cambio genético alguno.*

Este ejemplo pone de relieve el elemento de *deixis* —«señalar»— implícito en la clase de referencias que emplean las recetas genéticas. El gen de los descendientes de la mariposa dice, en efecto: pon tus huevos en algo que se parezca a *esto* (y un pequeño dedo señala ciegamente hacia fuera, hacia lo que sea que esté allí cuando el organismo «mire» hacia donde señala el dedo). En cuanto se ha comprendido el principio, es fácil verlo en

acción por todas partes, especialmente en procesos complejos de desarrollo que dependen de la «memoria celular». La mariposa no depositó simplemente su ADN en la hoja; depositó huevos, unas células que contienen toda la maquinaria de lectura y las materias primas iniciales para seguir las recetas del ADN. Esta maquinaria de lectura, a su vez, contiene información necesaria para elaborar el fenotipo de la descendencia, una información que no está codificada en los genes; éstos se limitan a «señalar» los ingredientes y a decirle a la maquinaria de lectura: usa *esto* y *aquello* para hacer y doblar la próxima proteína.[1] Si alteramos dichos elementos en el entorno inmediato del proceso de lectura de genes, podemos introducir un cambio en el resultado (igual que sucedía con el alterado hábito de elección de la hoja) y si resulta ser —igual que aquel hábito— un cambio que garantiza la recurrencia de la misma alteración en el entorno del proceso de lectura de genes, habremos producido una mutación del fenotipo (una mutación en el producto, el vehículo que se enfrenta a la selección natural) sin que comporte ninguna mutación en el genotipo (la receta). Los cocineros saben que un cambio sutil en la textura de la harina y el azúcar en países diferentes puede tener un efecto profundo sobre cómo les salen sus recetas favoritas. Siguen la receta al pie de la letra, buscan *eso que llaman harina aquí,* y obtienen como resultado un pastel distinto del que conocían. Pero si el nuevo pastel es un buen pastel, es posible que su receta sea copiada y seguida por muchos cocineros, y dé lugar a una línea de pasteles diferentes de sus antecesores y de sus parientes contemporáneos en el país original. (Confío en que los aficionados observarán los paralelismos que existen entre esta idea y la industria filosófica de la Tierra Gemela. Aquellos que no entiendan este paréntesis pueden considerarse afortunados por ello.)

La Madre Naturaleza no es «genecentrista». Quiero decir con ello que el proceso de selección natural no favorece la transmisión genética de la información cuando la misma información puede ser (más o menos) transmitida con igual fiabilidad, y de forma más barata, con la ayuda de alguna otra regularidad del mundo. Por un lado están las leyes de la física (la gravedad, etc.) y los elementos estables a largo plazo en el entorno en cuya perseverancia se puede «confiar» sin mucho riesgo (la salinidad del océano, la composición de la atmósfera, los colores de las cosas, que pue-

1. Los genes sí codifican la información para guiar la construcción de la maquinaria de lectura de la *próxima* generación, y para llenar la cocina de esa generación con las materias primas, pero tal como hemos visto hay otras fuentes que también pueden contribuir a dicha especificación.

den usarse como marcadores, etc.). Como esas condiciones son más o menos constantes, pueden estar tácitamente presupuestas en las recetas genéticas, sin necesidad de «mencionarlas». (Nótese que las mezclas para hacer pasteles que se venden en cajas a menudo prescriben diferentes temperaturas de cocción, o el añadido de más harina o más agua, para su preparación a gran altura, un ejemplo de variación que obliga a la receta a mencionar algo que de otro modo podría ahorrarse.)

Entre las regularidades que pueden presuponerse en las recetas genéticas están las que se transmiten de generación a generación por el aprendizaje social. Dichas regularidades sociales no son sino otros tantos casos de regularidades previsibles del entorno, pero adquieren una relevancia ulterior por el hecho de que ellas mismas pueden estar sujetas a selección (a diferencia de la gravedad, por ejemplo). Una vez que se ha determinado la vía de transmisión de información, y *los genes* pasan a «confiar en ella» para hacer parte del trabajo, ésta pasa a generar sus propias mejoras en el diseño, lo que ha dado lugar a una miríada de refinamientos que han ido puliendo los procesos de codificación, replicación, edición y transmisión del ADN a lo largo de eones. Los cambios genéticos que tienden a prolongar el contacto e interacción entre los progenitores y su descendencia, por ejemplo, pueden incrementar la Habilidad de la vía del aprendizaje social al darle más tiempo para operar, y más adelante se puede afinar la transmisión mediante el desarrollo evolutivo de marcadores de la atención (¡mira a mamá!). El camino se convierte primero en una carretera y luego en una autopista, un canal de transmisión de información especialmente *diseñado* por la selección natural para multiplicar la I+D en aquellos linajes que confían en él.

En aquellas especies en las que los padres conviven algún tiempo con su descendencia, hay una ancha avenida para dicha transmisión vertical pero no genética de información útil, o «tradición», como por ejemplo las preferencias en cuanto a la comida y al hábitat (Avital y Jablonka, 2000). Tal como hemos visto, la transmisión horizontal de un diseño *genéticamente transmitido,* la posibilidad de compartir genes útiles con organismos que no pertenecen a la propia descendencia o ascendencia, también ha estado presente desde los primeros días de la evolución y ha desempeñado un papel crucial en muchos de los avances más significativos de la evolución, pero estos casos parecen ser accidentes felices, no la vía prevista para la difusión de los diseños. La transmisión horizontal de información no genética es una innovación mucho más reciente en las formas de vida multicelular equipadas con sistemas perceptivos (en una palabra, los

animales). En ningún caso son más evidentes sus potencialidades que en nuestra especie, aunque no somos los únicos en disfrutar de sus beneficios. Un famoso estudio de los monos de una isla japonesa mostró que aprendían por imitación u observación el truco de limpiar el trigo de la playa a base de lanzar un puñado de trigo mezclado con arena al mar y luego recoger los granos que flotaban en la superficie, y hay razones para creer que las tecnologías de construcción de presas que los castores adultos transmiten a sus crías podrían incluir una parte importante de observación y aprendizaje por observación, si no de instrucción formal. También existen, como es habitual en biología, unos cuantos ejemplos intermedios para iluminar los contrastes. Las cabras montesas abren con su paso una red de caminos óptimos en su territorio, con lo que legan un entorno perfectamente preparado, tan pulcro como cualquier sistema de carreteras humano, no sólo a sus hijos y a sus nietos, sino a todas las criaturas que se mueven por la región. ¿Podemos considerarlo una transmisión cultural? Sí y no. La preservación de la uniformidad en la que se confía depende de la repetición de las acciones por parte de las cabras individuales, que deben ser capaces de ver lo que hacen las otras cabras. ¿Es eso imitación? ¿Qué es lo que se está *replicando?* Resulta difícil de decir.

Hay, sin embargo, una especie, el *Homo sapiens,* que ha convertido la transmisión cultural en una verdadera autopista de información, y ha generado grandes familias de entidades culturales que se ramifican en nuevas familias de familias, y ha transformado totalmente a sus miembros por medio del hábito culturalmente transmitido de inculcar tanta cultura como sea posible a los jóvenes, desde el momento en que sean capaces de absorberla. Esta innovación en la transmisión horizontal es tan revolucionaria que los primates que la realizan merecen un nuevo nombre. Si lo que buscamos es un término técnico podríamos llamarlos *euprimates* (superprimates). O podríamos usar el lenguaje corriente y llamarlos *personas.* Una persona es un homínido con un cerebro infectado, que se ha convertido en hospedador de millones de simbiontes culturales, y los principales factores que hacen posible esta transmisión son los sistemas simbióticos conocidos como lenguajes.

¿Qué fue primero, el lenguaje o la cultura? Como la mayoría de los dilemas del huevo y la gallina, éste sólo resulta paradójico cuando lo consideramos de una manera simplista. Es cierto que un lenguaje plenamente desarrollado no puede prosperar como institución entre los miembros de una especie hasta que no exista algo parecido a una comunidad, con normas, tradiciones, reconocimiento de individuos y funciones mutuamente

aceptadas. De modo que hay razones en favor de considerar que algún tipo de cultura precede —y debe preceder— al lenguaje. En las comunidades de chimpancés existen normas y tradiciones (en sentido amplio), reconocimiento de individuos y funciones mutuamente aceptadas (en sentido amplio), sin que esté presente el lenguaje, y también se advierte una modesta capacidad para la tradición cultural: tradiciones o «tecnologías» para romper cáscaras de nueces, cazar termitas o sacar agua de fuentes de difícil acceso. Disponen incluso de algunos protosímbolos: en al menos una comunidad de chimpancés, el picaro y lascivo gesto de acariciar una brizna de hierba por parte de un macho parece significar, a los ojos de una hembra, algo así como «¡Me pones a mil!» o «¿Estudias o trabajas?». Hay diferencias en las maneras de darse la mano durante los rituales de cortejo que parecen obedecer a reglas de transmisión cultural, no genética. Repasando nuestra propia historia evolutiva, hay pruebas (objeto aún de un encendido debate) de que el control de los homínidos sobre el fuego se podría remontar a un millón de años atrás, y seguramente debía ser una práctica de transmisión cultural (no genética, como las prácticas de cavado de nidos por parte de las avispas cavadoras), mientras que el lenguaje podría ser una innovación mucho más reciente: las estimaciones van desde cientos de miles de años atrás hasta sólo unas decenas de miles de años.

La cultura y la transmisión cultural pueden existir sin necesidad del lenguaje, y no son exclusivas de los homínidos, ni de los chimpancés, nuestros parientes vivos más próximos. Pero es el lenguaje el que abre las puertas para la transmisión cultural a gran escala que nos distingue de todas las demás especies. Parece que en nuestro planeta sólo ha evolucionado una cultura lingüística desarrollada (de momento). (Los Neandertales probablemente conocían el lenguaje, de modo que en cierto momento pudo haber en el planeta dos especies que lo usaban, pero, en caso de ser así, ambas debieron heredarlo probablemente de un ancestro común.) ¿Por qué no hay ninguna otra especie que haya descubierto esta maravillosa *suite* de las adaptaciones? Todos conocemos la lista de rasgos únicos del *Homo sapiens*: el control del fuego, la agricultura (aunque no hay que olvidar a las hormigas cultivadoras de hongos), las herramientas complejas, el lenguaje, la religión, la guerra (aunque debemos recordar también a las hormigas), el arte, la música, las lágrimas, la risa... ¿En qué orden surgieron estos rasgos peculiares, y por qué? Los hechos históricos se hallan muy alejados en el tiempo, pero no son inertes; han dejado rastros fósiles que pueden ser estudiados en la actualidad por los antropólogos, los arqueólogos, los genetistas evolucionistas, los lingüistas, etc. El común de-

nominador de todas las interpretaciones de los datos y de todos los debates actuales es el pensamiento darwinista, y no sólo a propósito de los genes. A veces ni siquiera se les dedica una mención. El lenguaje sólo ha evolucionado una vez, pero no han dejado de evolucionar *lenguas* distintas desde que el primer grupo conocedor del lenguaje se rompió en varios subgrupos, y, aunque ha habido sin duda respuestas genéticas al surgimiento del lenguaje (los cerebros han evolucionado anatómicamente para convertirse en mejores procesadores de palabras), es muy improbable que *alguna* de las diferencias que han surgido entre, por ejemplo, el finlandés y el chino, o el navajo y el tagalog se deban a las mínimas diferencias genéticas que pueden descubrirse (usando un sofisticado análisis estadístico) entre las poblaciones que tienen dichas lenguas como lengua materna. Cualquier niño puede aprender con la misma facilidad cualquier lengua humana a la que esté expuesto, por lo que sabemos. Eso indica que la evolución de las lenguas no está relacionada directamente con la evolución de los genes, a pesar de lo cual ha seguido patrones darwinianos: toda I+D tiene su precio, y cada nuevo diseño tiene que salir a cuenta de un modo u otro. Si persiste una u otra complejidad gramatical, por ejemplo, lo hace por alguna razón, puesto que *todo* cuanto existe en la biosfera está sometido a renovación, revisión o cancelación, en todo momento. Las costumbres y los hábitos están tan abocados a la extinción como puedan estarlo las especies, si no hay nada que los mantenga en pie. Las innovaciones sofisticadas —en el lenguaje o en otras prácticas humanas— no ocurren porque sí; lo hacen por alguna razón.

La pregunta es: ¿la razón de quién? Los abogados preguntan *Cui bono?*\ ¿a quién beneficia? Para responder adecuadamente a esta pregunta debemos realizar un atrevido salto con la imaginación, y sin la ayuda de ninguna pluma mágica. Tal como veremos, cuando demos el salto habrá una ruidosa multitud de observadores histéricos que nos advertirán de que no lo demos, que nos implorarán que giremos la espalda a esta peligrosa idea. El tema que estamos a punto de esbozar tiene un incomparable poder para irritar a los guardianes de la tradición y hacer que suba el volumen, aunque no la precisión, de sus críticas. Estamos a punto de considerar la noción de *meme,* un replicador cultural análogo al gen, y muchos de los que han considerado dicha idea han terminado por odiarla. Tratemos primero de comprenderla, sin embargo, y veamos si es realmente tan odiosa. Haré todo cuanto esté en mi mano para ofrecer una imagen vivida de las razones que hay detrás del odio para que no me acusen de dorar la píldora de una idea venenosa, y comenzaré a hacerlo desde ahora mismo.

Vemos cómo una hormiga escala laboriosamente el tallo de una hierba. ¿Por qué lo hace? ¿Por qué ha evolucionado tal comportamiento? ¿Qué beneficios obtiene la hormiga por hacerlo? No es ésa la pregunta que debemos plantear. La hormiga no recibe beneficio alguno. ¿Lo hace porque sí, entonces? En realidad, lo hace precisamente por eso: por un trematodo.* El cerebro de la hormiga ha sido invadido por un trematodo (*Dicrocoelium denáriticum*), perteneciente a una familia de pequeños gusanos parásitos que necesitan llegar hasta los intestinos de una oveja o una vaca para poder reproducirse. (Los salmones nadan contracorriente; estos gusanos parásitos hacen que las hormigas trepen por los tallos de hierba para aumentar la probabilidad de que las ingiera un rumiante.) El beneficio no es para las expectativas reproductivas de la hormiga, sino para las del trematodo.[2]

En *El gen egoísta* (1976), Richard Dawkins señala que también podemos concebir ciertos elementos culturales —a los que dio el nombre de *memes*— como parásitos. Tales elementos utilizan los cerebros humanos (en lugar de los estómagos de las ovejas) como hogares temporales y saltan de cerebro en cerebro para reproducirse. Igual que los tremátodos, han aprendido a negociar cada vez mejor este elaborado ciclo (debido a la competición entre los diferentes memes por el limitado espacio de los cerebros) y, también igual que los tremátodos, no tienen por qué saber nada acerca de cómo o por qué lo realizan. Son estructuras de información dotadas de un ingenioso diseño para explotar inconscientemente a los pensadores, lo que no significa que sean pensantes ellas mismas. No tienen sistema nervioso; ni siquiera tienen cuerpo, en el sentido ordinario. En realidad se parecen más a un virus que a un gusano (Dawkins, 1993), ya que viajan ligeras, sin necesidad de construirse un gran cuerpo con el que moverse. Básicamente, un virus no es más que una cadena de ácido nucleico (un gen) con una actitud. (También posee una especie de abrigo proteínico; un viroide es un gen aún más desnudo, desprovisto del abrigo.) De modo parecido, un meme es un paquete de información con una actitud: una receta o un manual de instrucciones para hacer algo cultural.

*** El original hace un juego de palabras entre *fluke* («golpe de suerte») y *lancet fluke* («trematodo») intraducibie al castellano. (N.** *del t.)*

2. Estrictamente hablando, para las expectativas reproductivas de los genes del trematodo (o de los genes del grupo del trematodo), pues tal como señalan Sober y Wilson (1998) al proponer el *D. dendriticum* como ejemplo de comportamiento altruista, el trematodo que se encarga de las labores de pilotaje en el cerebro de la hormiga es una especie de piloto kamikaze, pues muere sin ninguna posibilidad de pasar sus propios genes, en beneficio de sus cuasi clones (que se reproducen asexualmente) situados en otros lugares de la hormiga.

Así pues, los memes son análogos a los genes. ¿De qué está hecho un meme? Está hecho de información, que puede ser transmitida por cualquier medio físico. Los genes, que no son sino recetas genéticas, están escritos en el medio físico del ADN y en un único lenguaje canónico, el alfabeto de C, G, A y T, que se agrupa en tripletes para codificar los aminoácidos. Los memes, que son recetas culturales, dependen de modo parecido de uno u otro medio físico para seguir existiendo (no son mágicos), pero pueden saltar de un medio a otro, y traducirse de una lengua a otra, igual que... ¡las recetas! Una misma receta de pastel de chocolate puede ser conservada, transmitida y copiada sin importar si está escrita con tinta en un papel y en lengua inglesa, grabada en una cinta de vídeo en italiano o almacenada en una estructura diagramática de datos en el disco duro de un ordenador. Como la prueba de un bizcocho es el momento de comerlo, la probabilidad de que una receta consiga que se hagan réplicas de sus copias físicas depende (principalmente) del éxito que tenga el pastel. ¿Del éxito que tenga el pastel en qué? En conseguir un hospedador que haga otra copia de la receta y la pase a otros. *Cui bono?* Por regla general, los beneficiarios son quienes comen el pastel, y por ello conservan la receta como un tesoro, hacen copias y las transmiten a su vez; pero con independencia del beneficio que puedan sacar estos «anfitriones», si el pastel consigue de algún modo animarlos a que sigan transmitiendo la receta, ésta se beneficiará de la única manera que importa para las recetas: logrando que se hagan más copias suyas y que se prolongue su linaje. (Podemos imaginar, por ejemplo, que se tratara de una receta para hacer un pastel que resultara, en realidad, altamente tóxico, pero que contuviera un poderoso alucinógeno que despertara en las personas que lo comieran un deseo obsesivo e irresistible de hacer más copias de la receta y compartirlas con sus amigos.)

En el dominio de los memes, el beneficiario último, el beneficiario en términos del cual deben realizarse los cálculos finales de costes y beneficios, es el propio meme, no sus portadores. Esto no debe interpretarse como una tesis empírica radical, que descarte (por ejemplo) el papel de los agentes humanos individuales en el diseño, la apreciación y la contribución a la difusión y perduración de los elementos culturales. Mi propuesta es más bien adoptar una perspectiva o punto de vista que nos permita comparar una amplia variedad de tesis empíricas distintas, incluidas las tradicionales, y considerar las pruebas que hay a su favor desde una posición neutral que no prejuzgue dichas cuestiones. A primera vista, esta visión de la cultura puede parecer más siniestra que prometedora. Si *esto* es una forma de libertad, resulta sin duda extraña, y no parece en nada prefe-

rible a la libertad ignorante, aunque feliz, de la que disfruta el pájaro para volar allí donde quiera. A partir de una analogía con el trematodo, se nos invita a considerar un meme como un parásito que dirige a un organismo en aras de su propio beneficio replicador, aunque deberíamos recordar que dichos autoestopistas o simbiontes pueden clasificarse en tres categorías fundamentales: parásitos, cuya presencia reduce la competencia de su hospedador; comensales, cuya presencia es neutral (aunque, tal como nos recuerda la etimología, «comparten la misma mesa»); y mutualistas, cuya presencia aumenta la competencia tanto del hospedador como del invitado. Como dichas variedades se hallan distribuidas a lo largo de un continuo, no se pueden establecer límites muy precisos entre ellas; en qué punto cae el beneficio hasta cero o se convierte en perjuicio no es algo que se pueda medir directamente con ninguna prueba práctica, aunque podemos examinar las consecuencias de dichas variantes a través de modelos. Cabe esperar que los memes presenten también las tres variedades. Algunos memes seguramente promueven nuestra competencia y aumentan nuestras opciones de tener una descendencia numerosa (métodos de higiene, o para el cuidado de los niños o la preparación de la comida, por ejemplo), otros son neutrales —pero podrían ser buenos para nosotros en otros aspectos más importantes (la lectura y la escritura, la música y el arte, por ejemplo)— y otros memes son seguramente perjudiciales para nuestra competencia genética, aunque incluso éstos pueden ser buenos para nosotros en otros sentidos que nos interesan más (el ejemplo más obvio son las técnicas de control de la natalidad). Evidentemente, los memes que persistan serán aquellos que posean una mayor competencia como replicadores, sean cuales sean sus efectos sobre nuestra propia competencia, o siquiera sobre nuestro bienestar. En consecuencia, es un error *presumir* que la selección natural de un rasgo cultural se da siempre «por alguna causa», entendiendo por tal algo que *el hospedador* pueda percibir subjetivamente (tal vez por error) como un beneficio. Siempre cabe preguntar si los hospedadores, los agentes humanos que intervienen como *vectores,* perciben algún beneficio y (por esa misma razón, sea buena o mala) contribuyen a la preservación y replicación del elemento cultural en cuestión, pero debemos estar preparados para recibir la respuesta de que no es así. En otras palabras, debemos considerar como una posibilidad real la hipótesis de que los hospedadores humanos, individualmente o como grupo, no estemos interesados en cierto elemento cultural, tengamos reservas respecto a él o incluso estemos positivamente resueltos en su contra, y que sin embargo éste sea capaz de explotarnos como vectores. Tal como dijo George Williams:

Un meme puede promover la felicidad o la competencia de sus portadores dentro de una sociedad, pero también puede ser que no. Si puede transmitirse horizontalmente a un ritmo superior al que puede reproducirse su portador, la competencia de éste pasa a ser en buena medida irrelevante. El avance del tabaco deja un rastro de cadáveres que están tan muertos como las víctimas de una cepa de espiroquetas (Williams, 1988, pág. 438).

Quedan todavía muchas preguntas por responder respecto a los memes, y también muchas objeciones. ¿Podemos convertir el punto de vista de los memes en una ciencia propiamente dicha, la memética, o es «sólo» un vivido artificio para la imaginación, una herramienta o un juego filosófico, una metáfora que no puede tomarse en sentido literal? Es demasiado pronto para decirlo. La mayor parte de los argumentos que se han esgrimido contra una posible ciencia de la memética están mal planteados o son fruto de la desinformación, y desprenden un tufo inconfundible de hipocresía o desesperación. Ello es particularmente evidente cuando dichos argumentos los repiten personas que manifiestamente no los comprenden, puesto que replican con toda ingenuidad y sin darse cuenta pequeños errores que de algún modo se colaron en la línea germinal. Mi mala objeción favorita es la tesis de que la evolución cultural es «lamarckiana» y, por tanto, no puede ser «darwiniana», una letanía que se presenta en diversas variantes pésimamente formuladas, ninguna de las cuales se sostiene en pie.[3] Pero suena bien, ¿no es verdad? Suena como una objeción sofisticada capaz de darle donde más le duele a esa odiosa derecha ultradarwinista. (*¡Detengan a ese cuervo!*) La vanguardia de las investigaciones actuales puede terminar por convertirse en una disciplina sustantiva de la memética y demostrar que aquellos críticos estaban en un error. (*¡Cómete ésa, cuervo!*)* O puede ser que no. Todavía quedan obstáculos y objeciones importantes que superar. (Véanse las notas sobre lecturas com-

3. En pocas palabras, el lamarckismo es la herejía de la transmisión genética de los caracteres adquiridos, pero ¿de quién serían estos caracteres adquiridos, de los memes o de sus anfitriones? Sucede con frecuencia que los anfitriones transmiten sus parásitos adquiridos a su descendencia —ninguna herejía lamarckiana aquí—, y como los memes no poseen ninguna distinción entre la línea germinal y la línea somática, no hay distinción clara entre lo que es una mutación y lo que es una característica adquirida en un meme. Si la tesis de que la «evolución cultural es lamarckiana» significa alguna de estas dos cosas, no es ninguna objeción a la memética; si significa alguna otra cosa, todavía tiene que salir de detrás de la cortina de humo.

*** El original hace un juego de palabras con la expresión** *eat crow* **(literalmente «comer cuervo»), que significa «aceptar lo que antes se rechazaba». (N.** *del t.)*

plementarias incluidas al final del capítulo.) Tal como digo, es demasiado pronto para decirlo, pero eso no importa demasiado para nuestros propósitos, ya que la principal contribución que deben hacer los memes a esta exposición es en realidad «meramente» filosófica o conceptual, aunque no por ello menos valiosa: adoptar el punto de vista del meme nos permite *apreciar una posibilidad* que difícilmente nos tomaríamos en serio de otro modo. Tal como vimos en el capítulo 4, dedicado al libertarismo, existe un convencimiento ampliamente extendido entre los pensadores sobre la necesidad de librarnos de algún modo de nuestra herencia biológica para poder alcanzar la libertad relevante desde el punto de vista moral. Como no podemos recurrir a la levitación moral mágica y no podemos aprovecharnos de los quanta para que nos lleven más allá de nuestra biología, debemos buscar nuestra liberación en alguna otra parte. Richard Dawkins termina *El gen egoísta* con una resonante declaración:

> **Tenemos el poder de desafiar a los genes egoístas de nuestro nacimiento y, si es necesario, a los memes egoístas de nuestra educación [...]. Hemos sido construidos como máquinas genéticas e instruidos como máquinas memé-ticas, pero tenemos el poder de volvernos contra nuestros creadores. Sólo nosotros, en toda la Tierra, somos capaces de rebelarnos contra la tiranía de los replicadores egoístas (Dawkins, 1976, pág. 215).**

Pero, ¿cómo es posible que «nosotros» hagamos algo así? Dawkins no lo dice, pero pienso que el punto de vista del meme introduce precisamente las ideas que necesitamos para dar contenido a su proclama. Pero antes debemos dar unos cuantos pasos. El primero consiste simplemente en reconocer que el acceso a los memes —buenos, malos e indiferentes— tiene el efecto de abrir un mundo que de otro modo estaría cerrado a la imaginación de los seres humanos. La acción de nadar río arriba para desovar puede ser inteligente en muchos sentidos, pero el salmón no puede siquiera contemplar la posibilidad de abandonar su proyecto evolutivo y optar en cambio por pasar sus días estudiando geografía costera o esforzándose por aprender portugués. La creación de una panoplia de nuevos puntos de vista me parece el producto más extraordinario de la revolución euprimática. Mientras que todos los demás seres vivos están diseñados por la evolución para evaluar todas las opciones en relación con el *summum bonum* del éxito reproductivo, nosotros podemos sustituir este objetivo por miles de otros con la misma facilidad con la que el camaleón cambia de color. Los pájaros y los peces, e incluso otros mamíferos, son

en gran medida inmunes al *fanatismo,* una patología de origen cultural única en nuestra especie, pero a la que irónicamente nos hace vulnerables la propia cultura al darnos una *mente abierta* en cuanto a fines y medios, en un sentido que no es aplicable a ningún otro animal.

Cuando un agente o sistema intencional toma una decisión sobre cuál es el mejor curso de acción, tras considerar todos los factores, debemos preguntar desde qué perspectiva juzga su optimalidad. Una presunción que se da más o menos por defecto, al menos en el mundo occidental, y especialmente entre los economistas, es tratar al agente como si fuera una especie de punto o locus cartesiano de bienestar. ¿Qué me aporta eso a *mí?* Interés racional. Pero si bien es cierto que algo debe desempeñar el papel del yo, es decir, algo debe definir la respuesta a la pregunta *Cui bono?* para aquél que toma la decisión, no hay ninguna necesidad de adoptar dicha presunción, por más común que sea. El yo como beneficiario último puede estar indefinidamente distribuido. Puedo preocuparme por otros o por una estructura social superior, por ejemplo. No hay nada que me limite a un *yo* distinto de un *nosotros.* (Si uno se hace lo bastante pequeño, puede externalizarlo prácticamente todo.)

Cierta tradición hablaría aquí de ayuda «desinteresada», pero esto introduce más problemas de los que resuelve: la búsqueda del «auténtico» desinterés es una misión destinada al fracaso. Y debe fracasar no tanto porque no seamos ángeles (no somos ángeles, pero no es ése el problema), sino porque los criterios que definen el auténtico desinterés son en todo caso equívocos, tal como veremos. Es mejor concebir la capacidad humana para reformular su propio *summum bonum* como la posibilidad de extender el dominio del yo. El hecho de que mi objetivo sea convertirme en el número uno no se ve desmerecido en nada porque incluya en esta posición no sólo a mi propio cuerpo, sino a mi familia, a los Chicago Bulls, a Oxfam... lo que se quiera. Y hay una buena razón para concebir de este modo el yo: supongamos que soy un agente que participa en una negociación, o en el dilema del prisionero, o que me enfrento a una oferta coercitiva, o a un intento de extorsión. Mi problema no se resuelve, ni disminuye, o ni siquiera cambia en ningún sentido relevante, si el «yo» que estoy protegiendo es distinto del mío, si no estoy tratando de salvar mi propio pellejo, por así decirlo. Un extorsionador o un benefactor que sepa cuál es el objeto de mis cuidados está en posición de crear una situación que me afecte en lo que más me importa, sea lo que sea.

Estamos ya a las puertas de la sala de conciertos, pero todavía queda mucho por explorar. Debemos ver *de qué modo* llega a producir la evolu-

ción cultural, a veces en colaboración con la evolución biológica, las condiciones sociales que componen nuestra atmósfera conceptual, el aire que respiramos, cuando nos comportamos con el convencimiento de que muchas veces somos libres de tomar nuestras propias decisiones, *en un sentido moralmente relevante.*

LA DIVERSIDAD DE LAS EXPLICACIONES DARWINIANAS

Las ideas éticas, políticas, religiosas, científicas... todas estas ideas y las instituciones que las encarnan han surgido en un período biológico muy reciente, y no por arte de magia. La cultura no descendió un día sobre una banda de homínidos como una nube de gérmenes transportados por el aire. Para comprender cómo las ideas surgidas gracias a la cultura contribuyeron a ampliar nuestros yoes, debemos observar la estructura del entorno donde debieron actuar dichos agentes ancestrales. Al hacerlo, descubriremos una amplia y en gran medida inexplorada variedad de hipótesis darwinianas que deberemos contrastar en nuestra investigación de la historia que ha llevado a nuestra herencia cultural, y las razones que explican algunos de sus aspectos.

Un hábito cultural puede desaparecer de un día para otro cuando se produce un cambio en el entorno cultural, y esta desaparición puede tener a su vez efectos ulteriores sobre el entorno selectivo, lo cual supone un potente ciclo de retroalimentación que acelera el ritmo de la evolución, a menudo en direcciones que podemos terminar lamentando. Consideremos algunos ejemplos. La película de dibujos animados de Walt Disney *Bambi* salió en 1942 y en pocos años cambió las actitudes de los norteamericanos hacia la caza del ciervo (Cartmill, 1993). En la actualidad la población de ciervos se ha convertido en un grave problema de salud en algunas partes de Estados Unidos, donde ha llegado a provocar una epidemia menor del mal de Lyme, que es transmitido por las garrapatas de los ciervos a algunos seres humanos aficionados a pasear por el campo. Las latas de aluminio desplazaron en el curso de una sola generación a las tradicionales cestas sukuma de la cultura *masonzo,* en las costas del lago Victoria, en África:

> **Dichas cestas herméticas eran obra de las mujeres y servían en las celebraciones como recipientes para consumir grandes cantidades de** *pombe,* **una cerveza de mijo [...]. Las mujeres tejían las hojas de hierba desecadas con manganeso hasta convertirlas en cestas de diseños geométricos con un signi-**

ficado simbólico. No era siempre posible descubrir lo que significaban los diseños porque la llegada de los *mazabethi* —recipientes de aluminio llamados así por referencia a la reina Isabel y que habían sido introducidos a gran escala durante el mandato británico— había significado el fin de la cultura *masonzo.* Hablé con una anciana de un pequeño poblado que, después de más de treinta años, aún seguía indignada por la cuestión de los *mazabethi. [...] «Sisi wanawake,* nosotras, las mujeres, acostumbrábamos a tejer cestas sentadas en el suelo mientras charlábamos. No veo nada malo en eso. Cada mujer se esforzaba por hacer la cesta más bonita posible. Los *mazabethi* terminaron con todo eso» (Goldschmidt, 1996, pág. 39).

Más triste aún es el efecto que tuvo la introducción de las hachas de acero entre los indios Panare de Venezuela:

En el pasado, cuando se usaban hachas de piedra, era preciso que varios individuos se reunieran y trabajaran conjuntamente para cortar los árboles necesarios para hacer un nuevo jardín. Con la introducción del hacha de acero, sin embargo, un hombre solo puede despejar un jardín [...] la colaboración ya no es necesaria, ni tampoco muy frecuente (Milton, 1992, págs. 37-42).

Esa gente perdió su tradicional «red de interdependencia cooperativa» y ahora está perdiendo también buena parte del conocimiento que había reunido a lo largo de siglos, junto con la flora y la fauna de su mundo. A menudo desaparecen incluso sus lenguas, en sólo una o dos generaciones. ¿Podría ocurrimos algo así a nosotros? ¿Hay algún regalo de la ciencia o la tecnología que pueda tener un impacto parecido sobre nuestro medio cultural al que tuvieron aquellas sencillas hachas de acero en el suyo? ¿Por qué no? Nuestra cultura está hecha de las mismas cosas que la suya. (*¡Detengan a ese cuervo!...* sólo que ahora tal vez todos estemos de acuerdo en que tal vez haya buenas razonas para detenerlo.)

Estos ejemplos demuestran que los caracteres culturalmente mantenidos son muy volátiles y que se extinguen con facilidad en ciertas condiciones, lo cual resulta sin duda inquietante, aunque también es motivo de esperanza. Las perversiones culturales —como la tradición de la esclavitud o los abusos contra las mujeres— pueden evaporarse a veces en un período igualmente corto de tiempo, gracias a unos pocos cambios prácticos. No todos los caracteres culturales son igual de delicados. Un hábito culturalmente *impuesto* puede durar mucho más tiempo que el de su utilidad efectiva y persistir gracias a sanciones impuestas por los integrantes de la cultura, que pueden haber perdido de vista o tal vez apreciar sólo vaga-

mente el sentido original de su hábito convertido en tradición. Un tabú como el de no comer cerdo, por ejemplo, pudo tener un motivo perfectamente válido (virtual o no) en el momento de su establecimiento, un motivo que quizá dejó de existir hace mucho tiempo y que, sin embargo, ya no es necesario para el mantenimiento del tabú. Y si un carácter está fijado genéticamente, el lapso temporal entre la cesación de su *raison d'être* y su extinción efectiva puede medirse por cientos de generaciones. Nuestra afición al dulce, por usar un ejemplo muy gastado, era perfectamente razonable en un mundo de cazadores-recolectores, donde la conservación de la energía era una cuestión de vida o muerte. En el mundo actual, donde el azúcar está presente por todas partes en nuestro entorno, es una maldición que debemos superar con toda clase de contramedidas culturales. (Levantad las manos los deterministas genéticos que consideráis que eso es imposible... hmm, no veo ninguna mano.)

Hay muchas posibilidades de establecer interacciones complejas entre factores genéticos y culturales (así como con otros factores ambientales). Las simples diferencias de escala temporal garantizan eso por sí solas. Consideremos, por ejemplo, el siguiente estudio parcial de las posibles explicaciones darwinistas de la religión.[4] La religión es omnipresente dentro de la cultura humana y prospera a pesar de los considerables costes que supone. Cualquier fenómeno que excede aparentemente el terreno de lo funcional reclama una explicación. No nos maravillamos de que una criatura rebusque con insistencia entre la tierra con la nariz, porque suponemos que está buscando comida; si interrumpe regularmente su rastreo para dar una voltereta, en cambio, queremos saber por qué lo hace. ¿Qué beneficios supone la criatura (correcta o incorrectamente) que debe proporcionarle esta actividad suplementaria? Desde un punto de vista evolutivo, la religión parece ser una afición generalizada a dar las volteretas más elaboradas, y como tal reclama una explicación. No es que falten las hipótesis. La religión (o alguno de los rasgos de la religión) podría ser como:

> *El dinero:* **la religión es una innovación cultural bien diseñada cuya ubicuidad puede explicarse e incluso justificarse con facilidad: es un buen truco que previsiblemente volvería a ser descubierto una y otra vez, un caso de evolución social convergente. La sociedad se beneficia de ella. (Viene a ser como los rastros de feromonas que dejan los insectos sociales para coordinar las actividades de sus compañeros: su utilidad sólo puede comprenderse en el con-**

4. Los próximos párrafos están tomados, con algunas revisiones, de Dennett, 1997a.

texto del grupo, lo cual abre todas las cuestiones relacionadas con la selección grupal.)

Una estructura piramidal: **la religión es una estafa inteligentemente diseñada por una élite para aprovecharse de sus congéneres y que se ha venido transmitiendo (culturalmente). Sólo la élite se beneficia de ella.**

Una perla: **la religión es un bello subproducto de un mecanismo rígido y genéticamente controlado que responde a una irritación inevitable; el organismo se protege así de posibles daños internos.**

La glorieta de un pájaro glorieta: **la religión es el producto de algo análogo a una selección sexual desbocada, un proceso de elaboración de estrategias biológicas atrapado en una escalada de retroalimentación positiva.**

El tiritar: **esta agitación aparentemente inútil del cuerpo tiene en realidad una función positiva para el mantenimiento del equilibrio homeostático, ya que contribuye a elevar la temperatura corporal. El que tirita se beneficia de ello en la mayoría de los casos, aunque no en todos.**

Un estornudo: **los parásitos invasores han tomado el control del organismo y lo dirigen hacia destinos que les benefician a ellos, con independencia de cuáles sean sus efectos sobre el organismo, igual que hace el trematodo en el cerebro de la hormiga.**

La verdad acerca de la religión podría muy bien ser una amalgama de varias de estas hipótesis, u otras. Pero incluso aunque fuera así —sobre todo si fuera así— no conseguiremos hacernos una idea clara de por qué existe la religión hasta que hayamos distinguido claramente estas posibilidades y las hayamos comprobado todas. No todas apuntan en la misma dirección, aunque todas reflejan un pensamiento darwinista. Todas las hipótesis intentan explicar la religión tratando de descubrir algún beneficio, alguna función que justifique su coste, pero difieren ampliamente en la cuestión del *Cui bono?* ¿Es el grupo el beneficiario, la élite, el organismo individual? ¿Es un «efecto reina roja» en el que todos deben ir tan rápido como puedan tan sólo para seguir empatados? ¿Existe algún otro beneficiario evolutivo? Y ninguna de estas hipótesis invoca un «gen de la religión», aunque los genes tienen un papel principal en el establecimiento de algunas de estas hipotéticas precondiciones para ciertos aspectos de la religión.

Por supuesto, también *podría ser* que hubiera algo así como un gen para la religión. Por ejemplo, una especial tendencia hacia una «religiosidad» ferviente es un síntoma definitorio de ciertas clases de epilepsia, y se sabe que hay predisposiciones genéticas hacia la epilepsia. Podría ser que algunos entornos culturales —conjuntos de tradiciones, prácticas y ex-

pectativas— se convirtieran en factores de amplificación y configuración de ciertos extraños fenotipos, que tendieran a convertirse en chamanes, sacerdotes o profetas, cuyo mensaje sería el mensaje local que correspondiera según el caso (igual que como se aprende la lengua materna). Sólo en este sentido el «don de la profecía» podría «correr en la familia»: habría un gen para ello exactamente del mismo modo que hay genes para la miopía o la hipertensión. (Sí, sí, ya lo sé; «estrictamente hablando» no *hay* tal cosa como genes para la miopía y la hipertensión; esos supuestos genes no son más que predisposiciones hacia tales afecciones. *¡Detengan a ese cuervo!)* Aunque hubiera realmente algún gen para la religión, eso no sería más que una de las posibilidades darwinianas menos interesantes e informativas. Mucho más importante es la evolución (y el mantenimiento, frente a una posible extinción) de las condiciones que podían explicar su amplificación, y esto es algo que casi con toda certeza no está gobernado por los genes. Es evolución cultural.

Del mismo modo que propongo caricaturas del pensamiento darwinista, también podría poner en guardia contra otra de ellas, que llamo la falacia nudista. La revista *The American Sunbather* (algunos números de la cual cayeron en mis sudorosas manos cuando era joven) daba gran importancia, según recuerdo, al carácter esencialmente *natural* del desnudo. Era un retorno a nuestro pasado animal desnudo, una forma de conectar con «el modo como la Madre Naturaleza quería que fuéramos». Absurdo. Y no me refiero a lo de que la Madre Naturaleza quiera algo (yo mismo defiendo el uso de esta vivida forma de hablar para referirme a las razones virtuales que explican los diseños que la evolución descubre y suscribe). Lo que es absurdo es la idea de que aquello que la Madre Naturaleza quiere sea *ipso facto* bueno (para nosotros en este momento). No deje usted de quitarse la ropa cada vez que sienta el impulso de hacerlo, pero no cometa el error de suponer que al ponerse en una situación tan «natural» mejora de algún modo su situación en la vida. (En realidad, la ropa es tan natural para nuestra especie como la concha que toma prestada un cangrejo ermitaño, el cual sería más bien imprudente si fuera desnudo por ahí.) La miopía es natural, pero demos gracias por tener gafas. La Madre Naturaleza quería que comiéramos todas las cosas dulces que pudiéramos encontrar, pero eso no es una buena razón para seguirle la corriente a este instinto. Muchos de los elementos culturalmente evolucionados de la vida humana son eficientes correctivos de uno u otro «instinto» desfasado (Campbell, 1975), del mismo modo que otros elementos, tal como veremos, son correctivos de aquellos correctivos, y así sucesivamente. Los pro-

cesos darwinianos tienen su primer trampolín en la competición subyacente entre alelos dentro del genoma, pero en nuestra especie las adaptaciones han dejado el trampolín muy atrás.

BONITAS HERRAMIENTAS, PERO TODAVÍA FALTA USARLAS

> *Bajo la amable presión de las circunstancias, nuestras opiniones se revisan a sí mismas aprovechando que no prestamos atención. Les decimos con voz firme: «No, no me interesa cambiar en este momento». Pero no hay manera de tener quietas a las opiniones. No les importa si queremos mantenerlas o no; hacen lo que quieren.*
>
> NICHOLSON BAKER, *The Size of Thoughts*

> *En las últimas décadas, todo el mundo ha leído o visto un sinfín de libros dedicados a la cultura del narcisismo, del descreimiento, del deseo, de cualquier cosa. El argumento de estos libros es siempre el mismo: lo que crees que son tus bien fundadas creencias o preferencias resultan no ser otra cosa que un conjunto de reflejos implantados en ti por ciertas presunciones ocultas en tu «cultura». No eres escéptico con la religión porque no creas en la historia de Noé y el arca, sino porque formas parte de la cultura del descreimiento.*
>
> ADAM GOPNIK, *The New Yorker* (24 **de mayo de 1999)**

Antes de seguir adelante en este contexto tan cargado, es preciso exponer y desactivar una fuente ulterior de resistencia al pensamiento darwinista. Un profundo y persistente malentendido respecto al pensamiento darwinista pretende que siempre que se da una explicación evolutiva para un fenómeno humano, sea en términos de genes o de memes, se está negando que la gente piense. Algunas veces esta idea es un subproducto de la caricatura del determinismo genético, cuyos imaginarios seguidores tendrían la costumbre de decir: «La gente no piensa, simplemente tiene gran cantidad de instintos inconscientes». Pero también puede reconocerse la misma caricatura (a veces, debo admitirlo, una verdadera autocaricatura) en boca de algunos teóricos de la evolución cultural que dicen, en efecto: «Mis memes me obligaron a hacerlo», como si los memes (por ejemplo los memes del cálculo o de la física cuántica) pudieran hacer su

trabajo en los hospedadores humanos sin requerir de éstos ningún pensamiento. Los memes dependen de los cerebros humanos como sus nidos; los ríñones o los pulmones humanos no servirían como localización alternativa, porque los memes *dependen de* la capacidad de pensar de sus hospedadores. *Formar parte del pensamiento* es la forma que tiene un meme de ponerse a prueba y enfrentarse a la selección natural, del mismo modo que *conseguir que se realice la propia receta de proteínas y que el resultado salga al mundo* es la forma del gen de ponerse a prueba. Si los memes son herramientas de pensamiento (y eso es lo que son habitualmente los mejores de ellos), es preciso empuñarlas debidamente para que muestren sus efectos fenotípicos. Es preciso pensar.

Es cierto que un buen modelo darwinista sobre el pensamiento no tendrá el mismo aspecto que los modelos tradicionales. Debemos dejar atrás el viejo y erróneo modelo cartesiano de una *res cogitans* centralizada y no mecánica, una *cosa pensante* en sentido literal, encargada de hacer el trabajo espiritual relevante. Es preciso desmantelar el Teatro Cartesiano, el lugar imaginario situado en el centro del cerebro donde «todo confluye» ante la conciencia (y el pensamiento), y distribuir las funciones del pensamiento entre instancias menos fantásticas. En el próximo capítulo veremos con más detalle las consecuencias que trae consigo el hecho de que nuestras tareas de pensamiento sean delegadas a varios subcontratistas neurales semiindependientes en competencia entre sí, pero ninguna de ellas es que no siga siendo necesario pensar, y siempre que hay pensamiento *la gente hace las cosas por razones que son sus propias razones*.

Así pues, no es una cuestión de *memes contra razones*. Ni siquiera de *memes contra buenas razones*. Las explicaciones que pretenden dar sentido a tal o cual cosa a partir del razonamiento seguido por un agente pensante no quedan descartadas por una teoría darwinista seria. Nada más lejos. La única tesis acerca de las razones que contradice la memética es la tesis más bien incoherente de que las razones se las apañan de algún modo para existir sin ninguna clase de soporte biológico, colgadas de algún gancho celestial cartesiano. Una parodia servirá para poner en evidencia la falacia que hay detrás de todo ello: «La gente de Boeing ha caído en el ridículo error de pensar que *ha desarrollado* el diseño de sus aviones sobre la base de sólidos principios científicos y de ingeniería, y ha demostrado rigurosamente que los diseños son tal como deben ser, cuando *en realidad* la memética demuestra que todos esos elementos de diseño son memes que han sobrevivido y se han difundido entre los grupos sociales a

los que pertenecen dichos constructores de aviones». Sin duda es verdad que dichos memes han tenido éxito en tales círculos, pero eso no entra en competencia con la vieja explicación en términos de investigación y desarrollo racional debidamente planificada, organizada e implementada. No es sino un complemento de dicha explicación.

¿Por qué habría de pensar alguien de otro modo? Aparte de algunas confusiones ocasionales a este respecto por parte de algunos aprendices de darwinistas, y aparte de las inevitables caricaturas, hay una razón más interesante. A veces da la impresión de que los aprendices de memetistas niegan cualquier papel al pensamiento porque imitan la perspectiva que adoptan típicamente los genetistas de poblaciones, los cuales ignoran deliberadamente las actividades de los fenotipos cuyo éxito reproductivo diferencial determina el destino de los genes bajo estudio. Los genetistas de poblaciones tienden a obviar cualquier referencia a los cuerpos, las estructuras y los hechos del mundo real que de un modo u otro constituyen los factores de selección y se limitan a hablar de los efectos de tal o cual cambio hipotético sobre el acervo genético. Es como si los leones y los antílopes no vivieran realmente, sino que se limitaran a reproducirse o no, en función de los niveles de aptitud asignados a sus cuerpos. Imaginemos un torneo de tenis en el que los participantes simplemente se quitaran la ropa y fueran cuidadosamente examinados por parejas por un equipo de médicos y entrenadores deportivos que resolvieran por votación cuáles debían pasar a la siguiente ronda, hasta proclamar finalmente a un ganador. Los genetistas de poblaciones le verían todo el sentido a una práctica tan extraña como ésta, aunque reconocerían que los criterios de los jueces deberían basarse en el juego real, por lo que sería mejor dejar que los jugadores hicieran su parte y fueran sus enfrentamientos concretos los que decidieran a los ganadores. No dejarían de insistir, sin embargo, en que no es necesario mirarlos. El razonamiento estándar vendría a ser el siguiente:

> **Mientras el paso de un mecanismo al siguiente dé lugar a variaciones en la herencia, se producirán adaptaciones por selección natural. En cierto sentido no importa cuál sea en concreto el siguiente mecanismo. Si seleccionamos las alas largas en las moscas de la fruta y obtenemos alas largas, ¿a quién le importa el camino concreto que haya seguido este desarrollo? Si el trematodo ha evolucionado para sacrificar su vida para que el grupo termine en el hígado de una vaca, ¿a quién le importa qué (o si) piensa o siente cuando se cuela en el cerebro de la hormiga? (Sober y Wilson, 1998, pág. 193).**

De modo parecido, *podemos* ignorar la lucha entre memes que tiene lugar en los cerebros (después de todo, resulta terriblemente caótica y complicada), dar un paso atrás y limitarnos a registrar los ganadores y los perdedores, lo cual no debe hacernos olvidar que la competición sigue en pie. Hay pensamiento, y cómo sea este pensamiento afecta al éxito que tengan los diferentes memes.

Los algoritmos darwinianos sobre la evolución son *neutrales respecto al sustrato.* No se refieren a proteínas o al ADN, ni siquiera a la vida basada en el carbono; se refieren a los efectos de la replicación diferencial con mutación allí donde se produzca, sea cual sea el medio. Esto resulta especialmente importante cuando nos interesamos, como estamos a punto de hacer, por la evolución de la moral. Para apreciar esta neutralidad, consideremos una fantasía relacionada con otra creación específicamente humana: la música. Es muy probable que nosotros, los *H. sapiens,* tengamos alguna predisposición genética hacia la música. Pero sean cuales sean las probabilidades, supongamos que es así, por mor del siguiente experimento mental. Supongamos que nuestra pasión por la música, nuestra reacción ante la música, nuestro talento para la música, etc., son en parte el resultado de ciertos rasgos de diseño genéticamente transmitidos. Y supongamos también que esto nos distingue de unos «marcianos» inteligentes (una especie no humana pero culturalmente desarrollada y capaz de comunicarse) que desconocen por completo esta extravagante afición humana innata por la música. Un equipo de investigadores marcianos visita nuestro planeta. Uno de ellos se interesa, en un sentido intelectual, por la música de la Tierra, y se esfuerza por incorporar a sus propias proclividades y capacidades perceptivas todas las discriminaciones, preferencias, hábitos, etc., de un ser humano amante de la música. Mientras que un ser humano no necesita tomarse ninguna de estas molestias y es un amante de la música nato, para nuestro marciano imaginario se trata, sin duda, de un gusto adquirido. Pero supongamos que el marciano logra adquirirlo, gracias a un diligente esfuerzo de estudio y entrenamiento. Dejemos a un lado la cuestión (en último término aburrida) de si el marciano puede apreciar realmente la música «tal como la apreciamos los humanos». Consideremos en cambio la más interesante cuestión de cuáles son los criterios que distinguen la gran música de la música buena, mediocre o infumable.

¿Cuáles son los criterios que va a tener que apreciar el marciano si quiere convertirse en un crítico musical competente, por ejemplo? Esos son los criterios —íntimamente ligados, sin duda, a la peculiar historia ge-

nética del *H. sapiens,* pero describibles independientemente de ella— que más le interesaría descubrir a un teórico de la música darwinista. Supongamos que nuestro pionero marciano se lleva de vuelta a Marte la música terrestre y que otros marcianos adoptan este exótico pasatiempo y, siguiendo los pasos de su pionero, se imbuyen diligentemente de las actitudes y disposiciones requeridas (pero sólo a nivel cultural). Cuando *ellos* tocan, disfrutan o critican las obras de Mozart, la fuente de sus disposiciones será cultural, no genética, pero ¿qué más da? La cuestión de si alguien es un músico «natural» (diseñado genéticamente) o «artificial» (diseñado culturalmente) no tiene la menor importancia (desde ciertos puntos de vista relevantes). Las cuestiones relativas a las relaciones, las estructuras, las pautas que definen a Mozart, o a la música barroca, o a la música terrestre, son neutrales respecto al sustrato. Y si, como parece probable, la lista de éxitos marciana termina por incluir composiciones que nunca lograrían reunir a una audiencia en la Tierra, la explicación de las diferencias de sensibilidad entre marcianos y terrícolas que hay detrás de tales diferencias de gusto será neutral respecto a su origen genético o cultural. Ahora bien, si los marcianos fueran simplemente incapaces de adquirir estos gustos, nunca llegarían a exhibir los hábitos y las preferencias necesarias para perpetuar el fenómeno; los marcianos simplemente no tendrían oído para la música, no sería lo suyo. Pero si pudieran adquirir el gusto por la música, no importaría mucho en realidad cómo lo hubieran adquirido: la suma de las fuerzas de la naturaleza y de la crianza a lo largo de su desarrollo podría dar el mismo resultado aunque fuera por vías muy distintas, todas ellas darwinianas. Este experimento mental, a pesar de ser ciencia ficción, nos recuerda una importante verdad acerca de las diferencias entre los músicos humanos. Hay grandes diferencias entre aquellos que tienen un talento musical «natural» y aquellos que deben adquirirlo a base de internalizar grandes dosis de teoría. Es sin embargo algo próximo al racismo declarar que sólo los primeros son los músicos verdaderos, sólo los primeros tocan música *de verdad.* Sospecho que al final podremos identificar los genes «del» talento musical, pero la teoría musical es y debe ser neutral respecto a ellos.

Lo mismo debería decirse de la teoría que explique la moral. Debería ser neutral respecto a la cuestión de si nuestras actitudes, hábitos, preferencias y proclividades morales son producto de nuestros genes o de nuestra cultura. Desde un punto de vista empírico es importante saber hasta qué punto nacemos «de buena pasta», tal como ha dicho De Waal (1996) de los chimpancés, y hasta qué punto nacemos «torcidos» y teñe-

mos que confiar en la cultura para que nos haga rectos, tal como ha dicho Kant sobre nosotros: *Aus so krummem Holze, ais woraus der Mensch gemacht ist, kann nichts ganz Gerades gezimmert werden* [«De madera tan torcida como la de la humanidad no se ha podido hacer nunca nada derecho»]. La explicación de cómo surgió la moral y por qué es como es deberá ser en todo caso darwiniana. La interacción entre las vías de transmisión genética y cultural sólo puede examinarse desde una perspectiva neutral:

> **Incluso grupos genéticamente idénticos pueden diferir profundamente a nivel fenotípico a causa de mecanismos culturales, y esas diferencias pueden ser heredables en el único sentido relevante para el proceso de selección natural. El hecho de que la cultura pueda proporcionar por sí misma los ingredientes requeridos para el proceso de selección natural da a la cultura el estatus que los críticos del determinismo biológico no han dejado de subrayar (Sober y Wilson, 1998, pág. 336).**

Explicar por qué existe la música y por qué tiene las propiedades que tiene es un proyecto apenas esbozado. Explicar por qué existe la moral y por qué tiene las propiedades que tiene es otro proyecto, en el que tal vez se hayan realizado más progresos, y al que dedicaremos el próximo capítulo. Algunas de las intuiciones fundamentales proceden de los estudios ya tratados en el capítulo 5 en el campo de la teoría de juegos evolutiva. En años recientes ha surgido un grupo cada vez más multidisciplinar de investigadores que ha dedicado sus esfuerzos a explorar la evolución de la «cooperación», o el «altruismo», o el carácter «grupal», o la «virtud». Llámese sociobiología, psicología evolutiva, economía darwiniana, ciencia política, ética naturalizada o simplemente una rama interesante de la biología evolutiva, sus planteamientos revelan unas pautas que deben estar presentes en cualquier situación de conflicto de este tipo, vengan encarnadas por los genes, por los memes o por otras regularidades culturales. Recientemente han aparecido varios libros excelentes dedicados a repasar y explicar dichas investigaciones, y no trataré de realizar otro manual cuando otros lo han hecho ya tan bien (véase el apartado «Notas sobre fuentes y lecturas complementarias» incluido al final del próximo capítulo). En lugar de eso, adoptaré una perspectiva más amplia y trataré de dar algunas interpretaciones para orientar el tema hacia nuestros propósitos, así como algunos correctivos necesarios frente a la plaga de malentendidos que han venido sufriendo estas investigaciones.

Capítulo 6

Una teoría darwinista de la cultura humana nos permite esbozar un modelo explicativo capaz de justificar las principales diferencias entre nosotros y nuestros parientes más cercanos en el reino animal. La cultura es una innovación crucial dentro de la historia evolutiva. La cultura proporciona a una especie, el Homo sapiens, *nuevos temas sobre los que pensar, nuevas herramientas con las que pensar, y —puesto que los medios de la cultura abren la posibilidad de que haya replicadores culturales cuya propia aptitud sea independiente de nuestra aptitud genética— nuevas perspectivas desde las que pensar.*

Capítulo 7

La estabilidad de las condiciones sociales, las prácticas individuales y las actitudes en las que se funda nuestra agencia moral requiere un análisis y está cada vez más cerca de recibirlo, gracias a teóricos evolutivos que reconocen que la cultura debe obedecer también a las reglas de la evolución por selección natural. Frente a las sombrías advertencias de algunos críticos, este planteamiento no subvierte los ideales de la moral; más bien les proporciona un fundamento muy necesario.

NOTAS SOBRE FUENTES Y LECTURAS COMPLEMENTARIAS

Animal Traditions (2000), de Eytan Avital y Eva Jablonka, constituye una investigación fascinante sobre el tema escasamente estudiado de la tradición animal. Véase también mi reseña del libro (Dennett, en preparación b), que aparecerá en el *Journal of Evolutionary Biology,* así como la reseña de Matteo Mameli en *Biology and Philosophy,* vol. 17, n° 1 (2002).

Aquellos que quieran saber más acerca de la Tierra Gemela pueden consultar la antología de Andrew Pessin y Sanford Goldberg, *The Tivin Earth Chronicles* (1996), o mi artículo «Beyond Belief», en *La actitud intencional* (Dennett, 1987).

Respecto a los memes, véanse Blackmore, 1999; Aunger, 2000, 2002; Dennett, en preparición c; y un número especial de *The Monist* sobre la epidemiología de las ideas (Sperber, 2001). Más allá de *La peligrosa idea de Darwin* (Dennett, 1995) y de mis artículos en Aunger, 2000, y Sperber,

2001, he escrito sobre los memes en «The Evolution of Evaluators» (Dennett, 2001); una reseña de *Creation of the Sacred: Tracks of Biology in Early Religions* (Dennett, 1997a) de Walter Burkert; y un artículo general, «The New Replicators», en *Encyclopedia of Evolution*, de M. Pagels (comp.) (Dennett, 2002a).

Sobre la cuestión de por qué existe la religión puede encontrarse un excelente estudio en *Religion Explained: The Evolutionary Origins of Religious Thought* (2001), de Pascal Boyer.

El artículo de Gray y Jordan (2000) sobre la difusión del lenguaje en el Pacífico es un excelente trabajo acerca del empleo de métodos cladísticos para el análisis de la evolución lingüística. Mark Ridley (1995, pág. 258) trata la cuestión de los tremátodos y puede encontrarse una discusión más detallada en Sober y Wilson (1998). Cloak (1975) coincidió con Dawkins (1976) en la cuestión del *Cui bono?* de los elementos culturales: «El valor de supervivencia de una instrucción cultural coincide con su función; es el valor para la supervivencia / replicación de sí misma o de su réplica».

Para una discusión del error que supone confrontar las explicaciones darwinianas con las razones, véase mi comentario de «A Critique of Evolutionary Archaeology», de James L. Boone y Eric Alden Smith, en *Current Anthropology* (Dennett, 1998b).

Capítulo 7

La evolución de la agencia moral

Considero la moral una capacidad accidental producida, en su ilimitada estupidez, por un proceso biológico que normalmente se opone a la manifestación de tal capacidad.

GEORGE WILLIAMS, **en** *Zygon*

Si las comunidades de genes y células pueden desarrollar un sistema de reglas que les permita funcionar como unidades evolutivas, ¿por qué no pueden hacer lo mismo las comunidades de individuos? Si lo consiguen, entonces **los grupos serán como individuos,** *de acuerdo con la tesis que tratamos de probar.*

ELLIOTT SOBER **y** DAVID SLOAN WILSON, *Unto Others*

¿Es individualista o comunitarista la naturaleza? Habitualmente se piensa —sobre todo aquellos que temen cualquier invocación de consideraciones evolucionistas en el campo de la ética— que como el darwinismo presenta una «naturaleza con las manos manchadas de sangre», sólo puede subvertir o desacreditar nuestras aspiraciones éticas, nunca reforzarlas con nuevas intuiciones, nuevos fundamentos. Lo cual es falso.

BENEGOÍSMO

Debemos ir todos a una, o con toda certeza nos colgarán uno a uno.

BENJAMIN FRANKLIN **a John Hancock, durante la firma de la Declaración de Independencia, 4 de julio de 1776**

Esta exhortación de Ben Franklin todavía nos llama a través del tiempo, todavía parece que ondea al viento en rojo, blanco y azul, con aroma

de tarta de manzana, una frase sin duda inspiradora, elegante y noble, digna de que la dijera nuestro héroe, ¿verdad? Pero esperemos un momento. ¿No estaba apelando el viejo y astuto Ben a la prudencia cobarde e interesada de sus oyentes? Escuchad, cobardes, y permitidme que dirija vuestra atención hacia vuestra situación presente: unios a nosotros o morid. ¿Qué era eso, una apelación al altruismo y al sacrificio personal o una llamada a aquellos que sabían lo que más les convenía? Propongo que concedamos que no era, después de todo, una apelación a un *genuino* altruismo (más adelante consideraremos qué podría serlo, y si existe en alguna medida significativa), sino la expresión de algo que no deja de ser maravilloso: una llamada a un tipo de interés *previsor*, una clase de prudencia que tiende a perderse en el fragor de la competición debido a la célebre miopía de la evolución, que exige beneficios inmediatos para todas sus innovaciones. Propongo llamar a esta clase de cooperación previsora *benegoísmo,* en honor de Ben, pero también para sugerir que aunque sea una clase de egoísmo, es un egoísmo *del bueno.* Si no fuera por el feliz hallazgo de la elocuencia de Franklin, la hubiera llamado *euegoísmo.*

El altruismo genuino, o puro, es un concepto elusivo, un ideal que siempre parece evaporarse justo cuando estamos en posición de alcanzarlo. No queda claro en qué consistiría el genuino altruismo, y las paradojas acechan por todas partes a su alrededor. Imaginemos un mundo en el que sólo hubiera una persona altruista y todas las demás fueran egoístas. El altruista y uno de los egoístas se encuentran atrapados en una isla con un bote de remos en el que sólo hay espacio para uno. ¿Qué haría el altruista? ¿Debería ofrecerse voluntario para morir en la isla o sería mejor —más altruista— en su caso irse con el bote, dejando que el egoísta se las arreglara solo, para poder ayudar a unos cuantos tipos egoístas más en tierra firme? Un altruista no tiene por qué sacrificarse estúpidamente a cambio de nada: eso no es más que una estupidez. ¿Hasta dónde puede llegar un altruista en la explotación de otros para alcanzar sus propios fines altruistas? Consideremos por ejemplo la información de seguridad preceptiva que reciben los pasajeros en los aviones: si viaja con un niño, cuando bajen las mascarillas de oxígeno, póngase usted primero la máscara, y luego al niño. Parece que un padre puede seguir este consejo con la conciencia limpia, puesto que es probable (no hay nada seguro en esta vida) que si se cuida primero de sí mismo, estará en mejor posición de cuidar de su hijo, y el bienestar de su hijo es lo que más le preocupa. Eso le convierte en un altruista. Según Elliott Sober y David Sloan Wilson, en su libro *Unto Others: The Evolution and Psychology of Unselfish Behavior,* «la tesis

del altruismo, tal como nosotros la comprendemos, afirma que al menos parte del tiempo algunas personas consideran el bienestar de otros como un fin en sí mismo» (Sober y Wilson, 1998, pág. 228). Por supuesto, todo depende de lo que entendamos por un fin en sí mismo. Si usted, soñador egoísta, prefiere saborear en su imaginación las perspectivas de futuro de su hijo, si esta actividad le resulta más satisfactoria que ninguna otra y está dispuesto a dar todos los pasos necesarios para proteger a su hijo y preservar con ello la credibilidad de dichas fantasías paternas, entonces no se distingue usted en nada del avaro que arriesga su vida para salvar su cofre del tesoro de hundirse en el fondo del mar. Si cuando reflexiona sobre por qué lo está sacrificando todo por su hijo comete el error de comprometer su interés altruista por su hijo con un interés egoísta por su propia paz de espíritu, entonces no es usted ningún altruista. Usted sólo toma estas medidas para sentirse bien con usted mismo.

Y así sucesivamente, en una espiral de condiciones inalcanzables que estudiamos debidamente cada año en la clase de filosofía. Comienza cuando consideramos la célebre tesis de Sócrates (en el *Menón)* de que nadie desea el mal por sí mismo, una doctrina que resulta manifiestamente falsa hasta que se la respalda con el añadido de que nadie desea *a sabiendas* algo que es, *consideradas todas las circunstancias,* un mal en sí mismo. ¿Es cierta esta versión ponderada? ¿Es *imposible* o sólo *muy poco probable*? ¿Se trata simplemente de que alguien que deseara conscientemente cursos de acción que fueran, consideradas todas las circunstancias, malos en sí mismos probablemente no duraría lo bastante como para tener descendencia?

La muía es estéril a causa de los genes de sus padres, pero no porque haya heredado de ellos el «gen de la esterilidad», ya que no hay tal gen.[1]

1. Las muías tienen a un burro por padre y a una yegua por madre (por lo común; las muías que tienen a una burra por madre se llaman burdéganos); los burros tienen 62 cromosomas, y los caballos 64 (32 pares), y las muías tienen 63 cromosomas no emparejables. Son muy raros los casos de muías fértiles. Y hay condiciones bajo las cuales podría haber algo así como un gen de la esterilidad. Por ejemplo, puede haber un gen que en una dosis única (heterocigóticos: una copia de la madre o del padre, pero no de ambos) produjera grandes beneficios, tantos que persistiera a pesar de que aquellos con dobles dosis del gen (homocigóticos) fueran estériles. Se trata de una posibilidad que se pone sus propios límites, puesto que a medida que aumenta la proporción de aquellos que poseen una única copia del gen, aumenta también la probabilidad de que ambos padres tengan una única copia, y ambos la transmitan a su descendencia, con lo que crecería la proporción de descendientes estériles, un callejón sin salida para el gen. El ejemplo más conocido de este fenómeno harto familiar, la superioridad heterocigótica, es la resistencia a la malaria que produce la dosis única de un gen que en dosis doble provoca anemia celular falciforme.

La esterilidad es un callejón sin salida, el fin de una línea: no es algo que pueda transmitirse. ¿Es el altruista algo parecido a la muía, una conjunción más o menos azarosa de factores que es perfectamente *posible* pero incapaz en general de perpetuarse? Deberíamos tener presente que aunque las muías no tengan descendencia, sí proliferan en ciertos tiempos y lugares, gracias a efectos indirectos sobre otras especies (como los *Homo sapiens* pertenecientes a la Sociedad Protectora de las Muías británica, de la que obtuve algunos de estos detalles sobre las muías). En realidad, hay muchos caminos por los que la evolución puede dar lugar a poblaciones de organismos que a primera vista parece que deberían ser sistemáticamente descartados. Hay condiciones bajo las cuales ser altruista —o al menos benegoísta— no es un callejón sin salida genético ni cultural, y cada vez se amplía más la familia de modelos teóricos que exponen y clarifican dichas condiciones.

La variedad de modelos de teoría de juegos evolutiva que se han ido desarrollando a lo largo de las últimas décadas puede ordenarse, usando un poco el calzador, en algo así como un árbol genealógico de modelos a partir de una semilla original que da lugar a otros modelos y éstos a su vez a otros modelos y éstos a otros, y así sucesivamente hasta obtener un árbol que muestra —aproximadamente— dos tendencias entrelazadas: los modelos de partida son más simples que sus descendientes, la siguiente generación de modelos, y esta creciente complejidad de los modelos no trae consigo simplemente un aumento del realismo (los modelos reflejan cada vez más las complejidades del mundo real), sino también un aumento del optimismo. En los modelos simples, el altruismo parece condenado al fracaso. Aparte de algunas efímeras anomalías de la naturaleza, los altruistas parecen estar descartados por los principios fundamentales de la teoría evolucionista, son algo tan imposible como el móvil perpetuo. Es un mundo de todos contra todos, y los buenos tipos llegan *inevitablemente* los últimos. Más tarde, cuando se añaden algunos retoques para aumentar el realismo, aparece algo que va en la dirección del altruismo y que puede florecer en ciertas condiciones, y, a medida que añadimos nuevas capas de complejidad, parecen surgir todavía más variedades de cuasi altruismo, pseudoaltruismo o como queramos llamarlo. (Yo prefiero llamarlo benegoísmo.) Todo parece indicar que, a medida que nuestros modelos y teorías se vayan acercando a las complejidades del mundo real, terminaremos por llegar al genuino altruismo, como posibilidad real en el mundo real. ¿Es posible que esta perspectiva optimista no sea más que una ilusión? ¿Está condenado al fracaso este proyecto de partir desde abajo, viene a

ser algo así como querer construir una torre hasta la luna? Por este camino no se va a ninguna parte, dicen los escépticos antidarwinistas. No hace falta molestarse siquiera en intentarlo. Pero ¿no podría ser que fueran los escépticos quienes estuvieran equivocados, y mantuvieran una visión excesiva del altruismo que sólo resulta inaccesible partiendo desde abajo precisamente porque está sublimado, colgado del cielo?

En cualquier caso, todos los modelos muestran cuándo y cómo puede florecer el benegoísmo, y ninguno de los modelos desarrollados hasta el momento distingue entre el benegoísmo y el «genuino» altruismo, si es que puede definirse tal cosa. Todos muestran las condiciones bajo las cuales, en contra del viento dominante de la miopía de la evolución, ésta puede *diseñar* a los organismos para que cooperen, o, más precisamente, para que se comporten de modo que prefieran el bienestar a largo plazo del grupo a su propio bienestar individual inmediato.

La semilla de este árbol de modelos es el problema ilustrado por el dilema del prisionero. En esos modelos, la traición desempeña un papel parecido a la segunda ley de la termodinámica en física. Los físicos no dejan de recordarnos que las cosas se rompen, se desordenan, que *no* tienden a repararse a sí mismas a menos que intervenga algo especial, como por ejemplo un ser vivo, un agente local contra la entropía. De modo parecido, los economistas no dejan de recordarnos que no hay tal cosa como la comida gratis. Los evolucionistas, en la misma línea, nos recuerdan que siempre terminarán por surgir los oportunistas, y cuando lo hagan pronto ganarán las competiciones locales por la reproducción, a menos que haya algo que lo impida. Sea cual sea el juego local, y sean cuales sean los costes y beneficios para el *grupo* (la población local que debe compartir el espacio, los recursos y los riesgos), si es posible compartir los beneficios de la acción colectiva sin pagar la parte que a uno le corresponde de los costes (lo que uno debe, podría decirse), entonces aquellos que sigan la vía egoísta prosperarán más que los otros. Es tan fácil como hacer una resta: los *beneficios netos* (beneficios menos deudas) tienen que ser inferiores a los *beneficios brutos,* que son los que disfruta el oportunista por definición. Todo esto debe ser cierto a menos que haya condiciones de algún tipo que lo impidan. Comencemos con una población uniforme de felices cooperadores (todos tienen el gen de la cooperación, para hacerlo sencillo). Hay que suponer que en general sus hijos son como ellos, pero ¿qué ocurre si en una generación surge un muíante oportunista? El oportunista obtiene como mínimo los mismos resultados que los cooperadores (pues no paga lo que debe) y, por lo tanto, tiene una descendencia de oportunistas superior a la

media. Muy pronto habrá una tribu cada vez más grande de oportunistas y, con independencia de lo bien o mal que le vayan las cosas al grupo (es probable que le vayan peor, con la carga que suponen todos esos oportunistas), nadie en el grupo prosperará más que los oportunistas, que gradualmente terminarán por dominar el grupo.

Naturalmente, algo debe intervenir para impedir esta lamentable degradación. Podemos imaginar, si queremos, que los oportunistas tienden a ser estériles, o infanticidas. ¡Qué bien les iría eso a los cooperadores! También podemos imaginar que Zeus disfruta lanzando rayos contra los oportunistas, y que mantiene su número controlado (¡gracias a Dios!) mediante la práctica de este deporte. Dejando a un lado las fantasías fáciles, podemos preguntarnos qué tipo de *producto natural de la evolución* podría tener el efecto de bloquear de manera sistemática la toma del poder por parte de los oportunistas, que debemos suponer como la tendencia por defecto. Tal como hemos visto, este problema ya surgió en los primeros días de la vida en este planeta, en el conflicto intragenómico entre buenos genes y genes parásitos oportunistas, y se resolvió mediante la evolución de mecanismos de compensación que mantenían controlados a los oportunistas. Por supuesto, los problemas que se suscitaran a aquel nivel temprano y microscópico estaban fuera del alcance de Darwin, pero él mismo reconoció el problema en el caso de los insectos sociales, cuya devoción extrema al grupo era todo un reto para la teoría de la selección natural. William Hamilton mostró en sus famosos artículos sobre «selección por familias» cómo podían haber desarrollado los insectos sociales (y otras especies altamente sociales) dichos patrones de instinto cooperativo, y Richard Dawkins reformuló el modelo de Hamilton desde la perspectiva del gen egoísta. En el caso extremo de dicho comportamiento autosacrificial nos vemos obligados a bajar al nivel del gen para encontrar respuesta a la pregunta del *Cui bono?,* ya que, de acuerdo con la gráfica expresión de Sterelny y Griffiths: «Tal vez sea astuto por parte de un petirrojo optar por no poner todos los huevos que puede, pero una abeja que pica a un intruso, a un coste cierto para su vida, no puede estar guardándose nada en la manga» (Sterelny y Griffiths, 1999, pág. 157).

Los primeros modelos suponían, en beneficio de la simplicidad, un único gen para la «cooperación» y otro gen alternativo para la «traición», y se consideraba que dichos genes operaban de manera determinista *en el nivel biológico del comportamiento.* (Recuérdese: esto no tiene nada que ver con el determinismo o el indeterminismo físico y sí en cambio con el *diseño.* En dichos modelos se estipula que los organismos individuales de-

ben ser perros viejos incapaces de aprender nuevos trucos y que son unos colaboradores o unos traidores para toda la vida.) Este planteamiento no resulta una simplificación excesiva cuando hablamos de insectos, cuyas rutinas de comportamiento son relativamente simples y tropísticas (o *sphexish* [sphexístico] por usar el término acuñado por Douglas Hofstadter, en honor de la avispa *Sphex*), aunque incluso los insectos sociales pueden mostrarse sorprendentemente flexibles en ciertas condiciones, y pasar de la noche a la mañana de ser un zángano a convertirse en un trabajador cuando las circunstancias de la colonia exigen una recolocación, por ejemplo.

Los modelos muestran que los traidores tienden a prosperar, aunque también pueden contaminar sus propios nidos: a medida que aumenta la proporción de oportunistas, éstos tienden a encontrarse entre sí con más frecuencia, lo que genera caros episodios de defección mutua, y ya no hay bastantes cooperadores explotables cerca para marcar la diferencia. De este modo los cooperadores inician un tímido contraataque, pero sólo hasta que hay bastantes como para que valga la pena aprovecharse de ellos, en cuyo momento los oportunistas comienzan a prosperar otra vez. Pero los modelos también mostraban algunos efectos extraños que llevaban a equilibrios que no se ajustaban a nuestras expectativas y planteaban, por lo tanto, la duda de que al menos parte del comportamiento de los modelos fuera artificial, un subproducto de las simplificaciones más que un reflejo de algo que pudiera tener lugar en el mundo real. (Para un tratamiento lúcido de la cuestión véase Skyrms, 1996.) Esto se parece al descubrimiento místico de que, de acuerdo con nuestro modelo aerodinámico, los abejorros no pueden volar. Algo debe fallar en el modelo, puesto que por ahí va zumbando un abejorro. El modelo debe ser excesivamente simple, debe dejar fuera una complicación que resulta crucial para el manifiesto éxito del abejorro. Una de las simplificaciones de dichos modelos de teoría de juegos evolutiva era su carácter excesivamente abstracto. Los individuos eran meramente miembros de un conjunto, tomados por parejas al azar para que realizaran interacciones que determinaban su destino en la siguiente fase, sin preocuparse por sus ubicaciones espaciales relativas en el mundo. Era como si los organismos individuales vivieran en Internet, con la misma probabilidad de interactuar con alguien que está al otro lado del mundo como con el vecino de al lado. (En realidad, la interacción en Internet está muy ordenada; ciertas personas están mucho más «alejadas» entre sí —es más difícil acceder a ellas— que otras, de modo que esos modelos simplificarían gravemente incluso la «aldea glo-

bal» de la Red.) Una segunda ola de modelos impuso una espacialidad simplificada al vincular la probabilidad de los encuentros a un factor de «viscosidad» (cuanto mayor era la viscosidad del espacio imaginario, más probable era que una persona interactuara con alguien que tuviera una dirección cercana), y este sencillo cambio introdujo nuevas oportunidades para la evolución de la cooperación, al tiempo que eliminaba aquellos incómodos equilibrios. Resulta que la *vecindad matea,* una gran diferencia. (Las irrupciones son lo que hace la vida interesante.) La vecindad hace más probable que uno interactúe con aquellos que más se le parecen, con lo que obtiene mejores resultados de cualesquiera comportamientos cooperativos que adopte, puesto que es más probable que reciban una respuesta recíproca.

Si hacemos todavía un poco más sofisticados a los agentes individuales y les permitimos cierto margen para *elegir* con quién quieren interactuar (simplemente les permitimos negarse a jugar en ciertas condiciones, para empezar), el sencillo espacio que habitan todos (no muy diferente del plano del mundo Vida) comienza a adquirir cierta estructura: comienzan a formarse congregaciones de agentes que se comportan de modo parecido, lo que supone la creación de grupos con caracteres distintivos. Los cooperadores tienden a buscar otros cooperadores y los traidores terminan por tener que asociarse con otros traidores. Todo esto resulta muy sugerente, por supuesto, pero todavía estamos muy lejos del altruismo. Por ejemplo, ¿no rechazaría un genuino altruista la egoísta política de buscar otros altruistas para agruparse? ¿No debería el genuino altruista apartarse de su camino para ser el único altruista en medio de un grupo egoísta? Allí es donde más lo necesitan, podría decir alguien, y no en medio de sus compañeros altruistas. ¡Qué meramente benegoísta por su parte hacer lo que hace! Por otro lado, los agentes de estos modelos siguen siendo perros viejos de ideas más bien fijas, máquinas de situación-acción con unos cuantos interruptores prefijados que determinan sus «elecciones» en cualquier encuentro mediante la aplicación de una sencilla regla. Un vivido recordatorio de lo simples que son los agentes en estos modelos es que las estrategias de autosegregación y ostracismo que emergen de dichos modelos ya fueron explotadas a nivel macromolecular durante el conflicto intragenómico de la era procariota. Un modelo que no tiene necesidad de distinguir entre una macromolécula y un ciudadano humano adulto resulta escandalosamente abstracto.

Si hacemos todavía más flexibles a los agentes, más plásticos, y les damos la posibilidad de aprender de su experiencia, y reajustar sus reglas de

nacimiento en función de los encuentros que han tenido, las cosas se ponen aún más interesantes. La inevitabilidad —nótese el término— de que un grupo sea tomado por los oportunistas ha dependido hasta ahora de la presunción de que nadie se daría cuenta; ninguno de los individuos tendría capacidad para advertir lo que ocurría, dar la alarma, lamentarlo, proponer sanciones, formar grupos de vigilancia, marcar o castigar a los oportunistas que hubiera entre ellos. Tan pronto como incorporamos algunas versiones simplificadas de esta capacidad de reacción, vemos como se multiplica la complejidad de los modelos. Situaciones muy negativas que antes parecían inevitables desaparecen con sólo un poco de prevención, gracias a un uso oportuno y bien dirigido de la información por parte de los miembros del grupo. Los tipos benegoístas tienen ahora una razón para castigar a los «altruistas» demasiado puros —los inocentones o los blandengues que siempre se dejan explotar por los oportunistas—, puesto que son estos primos quienes permiten que los oportunistas se salgan con la suya. De este modo resultarán favorecidas todas las mutaciones que permitan a los benegoístas distinguirse de los primos, pero todos los oportunistas o los primos que puedan hacerse pasar por benegoístas tenderán a prosperar, hasta la siguiente fase de la carrera armamentística. El desarrollo de la capacidad del grupo para controlar a sus miembros mediante la adopción de disposiciones para castigar las transgresiones (de cualesquiera políticas que mantengan) abre las puertas a la evolución *social* o *cultural* de toda clase de normas locales. En un artículo ya clásico sobre la evolución cultural, Rob Boyd y Peter Richerson muestran que si el coste de castigar es *relativamente* bajo —algo que puede garantizarse casi siempre que surge la práctica de castigar a aquellos que no castigan lo bastante—, se crea una máquina generadora de conformismo grupal de alcance y poder aparentemente ilimitado. El título del artículo lo dice todo: «Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups» [«El castigo hace posible la evolución de la cooperación (o de cualquier otra cosa) en grupos de cierto tamaño»] (Boyd y Richerson, 1992).

Hasta aquí, pues, nuestro relato evolucionista ha sugerido una serie de condiciones que podrían habernos llevado, sin necesidad de ganchos colgados del cielo ni de otros milagros parecidos, a una prudente disposición en favor de la cooperación, reforzada por la disposición que compartimos con los demás ciudadanos a «castigar» a aquellos que no cooperan, pero todavía se trata de una no agresión mutuamente impuesta de carácter más bien frío y robótico. Tal como dice Alian Gibbard:

Las propensiones naturales humanas fueron el producto de algo que sería absurdo valorar en sí mismo, a saber, la voluntad de multiplicar los propios genes en las generaciones posteriores. Sin embargo, los modelos de coordinación que ayudaron a nuestros antepasados a transmitir sus genes para formarnos a nosotros sí que merecen la pena, por otras razones. Las fuerzas darwinianas han dado forma a las preocupaciones y los sentimientos que conocemos, y algunos de éstos son en gran medida morales (Gibbard, 1990, pág. 327).

En gran medida morales, pero no puramente morales. No hay el menor indicio de una voluntad de tratar el bienestar de otros como un fin en sí mismo, por ejemplo. Y así es probablemente como debe ser, puesto que todavía nos falta incluir algo específicamente humano en nuestros modelos, y una de nuestras intuiciones iniciales menos controvertidas respecto a la moral es que aunque los animales no humanos puedan estar hechos de «buena pasta», tal como dice Frans de Waal, no son todavía «el animal moral», tal como dice Robert Wright. Sin embargo, cabe considerar esta clase de estructura, social capaz de autoperpetuarse como una precondición necesaria para el florecimiento a largo plazo de agentes genuinamente altruistas, y en este sentido resulta alentador ver lo poco que debe presuponerse para verla evolucionar y sostenerse: el carácter relativamente simple y rígido de la capacidad de discriminación entre oportunistas y buenos ciudadanos, así como de la disposición a «castigar», indican que al menos *este* aspecto de la cultura podría ser anterior al lenguaje, a las convenciones y a las ceremonias. No estamos hablando de un juicio público y con jurado; estamos hablando de una inclinación irreflexiva y «brutal» a canalizar cierta agresividad hacia aquellos miembros del grupo que se han identificado como violadores de normas. Sería razonable buscar pruebas de esta clase de mantenimiento de «costumbres» locales duraderas en manadas de lobos o bandas de monos o simios, por ejemplo. Encontremos o no manifestaciones claras de este estadio de desarrollo previo a la cultura plenamente humana en alguna otra especie, la idea proporciona un cierto alivio frente al escepticismo: nos ofrece una *posible* «Historia de así fue» sobre la transición gradual desde unos animales meramente sociales a la manera de las abejas o las hormigas hasta unos animales con cierto gusto por la transmisión y la enseñanza de la cultura, interesados en los matices de la aprobación o la desaprobación, dispuestos a alistarse en pelotones temporales de castigo, inclinados a preferir el confort de la aceptación a la amenaza de la censura del grupo. Y gracias a esta transición los grupos se convierten en depositarios eficientes del recién descubierto «conocimien-

to», sin necesidad de esperar a que cada nuevo buen truco deba evolucionar y difundirse por vía genética hasta su fijación entre la población, puesto que el conformismo grupal puede garantizar una difusión mucho más rápida. Bien merece pagar el precio de una cierta vulnerabilidad ante cosas como los mitos, ante los descubrimientos *erróneos* locales que a pesar de todo venden igual de bien gracias al estructurado conformismo del grupo, a cambio de acceder a este ritmo más ágil de descubrimientos.

SER BUENO PARA PARECER BUENO

jesús se acerca. ¡Haz como si estuvieras ocupado!

Pegatina de coche

La conciencia es la voz interior que nos avisa de que alguien podría estar mirando.

H. L. MENCKEN, *Prejudices*

El espectro de la traición se cierne sobre todos nosotros; es el pecado original de la evolución, que nos llama con su reflexión perennemente tentadora: ¿cómo puede *no* ser racional traicionar en esta situación? Si el otro traiciona (o si «todo el mundo lo hace»), eres un primo si no lo haces, y si el otro no lo hace, te sales con la tuya como un fuera de la ley. Y si todo el mundo es consciente de esto, ¿cómo es posible que alguien colabore? Si las retribuciones que se obtienen con ello son a corto plazo, ¿cómo puede ignorarlas la evolución? Y, teniendo en cuenta que la vida es corta, ¿cómo podemos ignorar nosotros mismos estas retribuciones? El miedo al castigo y la tendencia a la aceptación nos permitirán resistir en los casos más fáciles, con sólo cambiar la retribución esperada. Tal como los pensadores han reconocido desde hace siglos, no cuesta mucho comprender por qué es racional cooperar cuando nos mira el Gran Hermano. Cualquier sociedad que tuviera la fortuna suficiente de desarrollar una fe en un Dios omnipresente y vigilante —del que cupiera esperar que impusiera un castigo en una vida ulterior capaz de compensar cualquier beneficio presente— sería una sociedad poblada por ciudadanos que cumplirían con toda fiabilidad los mandamientos de Dios, incluso cuando no estuvieran a la vista de los demás ciudadanos. Nótese que para el surgimiento y la difusión de este mito no tiene por qué haber ningún autor inteligente

que comprenda esta razón de fondo, del mismo modo que no tuvo por qué haber ninguna instancia inteligente que promulgara las estrategias que llevaron a garantizar el sometimiento a las normas de los genes rivales gracias a la meiosis. Los seres humanos podrían ser los beneficiarios inconscientes de esta adaptación grupal sin que nadie fuera consciente de la razón virtual que hay detrás de ella. Pero tal como han insistido los críticos desde Nietzsche, la «moral» basada en el temor de Dios no es tan noble ni tan estable como desearíamos. ¿Qué ocurriría en una sociedad en la que comenzara a venirse abajo todo este andamiaje, o en la que nunca hubiera existido? ¿No habría ninguna forma de que sus miembros desarrollaran unos firmes hábitos cooperativos?

¿Qué pasa con los casos más difíciles, en los que uno puede estar prácticamente seguro de que no se descubrirá el engaño? En esos casos la voz de la tentación habla con alarmante racionalidad: *¡nadie lo sabrá nunca, y piensa en lo que puedes ganar!* Tan pronto como entramos en un mundo donde la capacidad de decisión debe lidiar con tentaciones importantes y con los ilimitados meandros de la reflexión que pueden acompañar nuestras batallas contra la tentación, dejamos atrás la libertad de los pájaros y comenzamos a explorar el problemático territorio de la libertad humana, la única variante que posee un peso moral. La tradición pone la carga de todo ese peso moral sobre las espaldas de un funcionario imaginario, el alma inmortal, inmaterial, hacedora de milagros, pero si miramos más de cerca los antecedentes de nuestros sistemas humanos de control, podemos ver los pasos que han llevado al diseño de esa alma y ver cómo funcionan algunas de sus partes.

Según Salustio, Catón era un hombre de gran nobleza: *Esse quam videri bonus malebat* [«Prefería ser bueno antes que parecerlo»]. Si Robert Frank está en lo cierto, Catón era una de esas escasas almas adelantadas que han logrado darle la vuelta a la política que nos hizo morales en un principio: *Malo esse bonus ut videar* [«Prefiero ser bueno *para parecer* bueno»]. En *Passions within Reason: The Strategic Role of the Emotions,* Frank sostiene que el siguiente paso en la evolución de la libertad se produjo cuando nuestros antepasados se enfrentaron por primera vez a lo que llama *problemas de compromiso* y aprendieron a resolverlos. Un problema de compromiso «surge cuando el interés de una persona consiste en adoptar un compromiso vinculante para comportarse de un modo que más adelante parecerá contrario a su interés» (Frank, 1988, pág. 47). Ya hemos encontrado la estructura básica de un problema de compromiso en el dilema del prisionero: el destino evolutivo de los cooperadores y los

traidores se ve profundamente afectado por la presencia o ausencia de falsos cooperadores, o faroleros. Ello genera una presión selectiva en favor de la detección de faroleros y pone en marcha una carrera armamentista de estrategias de desenmascaramiento y ocultación. Cuando los flexibles sistemas de control de los agentes humanos se ajustan a las razones virtuales de esta competición, el tempo se acelera y la cuestión pasa de ser impersonal (¿qué agentes tendrán mayor éxito bajo las condiciones actuales, los cooperadores o los traidores?) a ser personal (¿qué debería hacer *yo* en estas condiciones, cooperar o traicionar?). Cuando la evolución logra crear agentes capaces de aprender, reflexionar y considerar racionalmente sus acciones, pone a esos agentes ante una nueva versión del problema del compromiso: ¿cómo comprometerse a algo *y convencer a otros de que lo han hecho*? Llevar una gorra que diga «soy un cooperador» no nos llevará muy lejos en un mundo de seres racionales a la caza de tramposos. Según Frank, en el curso de la evolución «aprendimos» a aplicar nuestras emociones a la tarea de impedir que fuéramos demasiado racionales, y a crearnos una reputación —lo que es igual de importante— de no ser demasiado racionales. Según Frank, es nuestro exceso indeseado de racionalidad miope o local lo que nos hace tan vulnerables a las tentaciones o las amenazas, tan vulnerables a «ofertas que no podemos rechazar», como dice el Padrino. Parte de lo que supone convertirse en un agente verdaderamente responsable, en un buen ciudadano, es convertirse en alguien en quien *se puede confiar* que será relativamente impermeable a tales ofertas.

En primer lugar, ¿por qué habríamos de querer una reputación como ésta? Pues bien, porque en caso de tenerla la mafia nos dejará tranquilos, pues calculará que sus ofertas coercitivas no tendrán probablemente éxito con nosotros, de modo que: ¿por qué malgastar una buena cabeza de caballo? Más importante aún: nuestra reputación nos hará más susceptibles de ser elegidos por nuestros compañeros de grupo, que son muy conscientes del riesgo que supone ser tomado por sorpresa por un traidor y que están a la búsqueda de alguien en quien depositar su confianza por considerarle capaz de resistir a la tentación. Hemos señalado ya en la sección anterior que los cooperadores tienden a juntarse con los cooperadores, y los traidores con los traidores. «Los problemas de compromiso son muy frecuentes, y hay toda clase de ventajas materiales esperando a aquellos cooperadores que sean capaces de encontrar a sus semejantes», observa Frank (1988, pág. 249); las ventajas de ser un cooperador en un grupo de cooperadores han quedado demostradas en un sinfín de modelos evolutivos. Si tenemos la fortuna de encontrarnos en un grupo de cooperadores, ¿debe-

mos considerar que ha sido pura suerte? No si el grupo tiene que pasar un examen de ingreso. Pero en tal caso, ¿debemos considerar que ha sido pura suerte que poseyéramos el talento para la cooperación que nos ha permitido pasar el examen? Tal vez, pero tener la suerte de poseer un talento es mejor que tener suerte a secas. (Más adelante volveré sobre el tema de la suerte.)

Es benegoísta querer una reputación impecable, ¿pero cómo podemos ganárnosla? Hablar es fácil, por lo que todo el mundo estará dispuesto a jurar sobre una pila de Biblias que nunca sería capaz de traicionar a nadie. A menos que haya alguna otra forma de distinguir a los cooperadores de los traidores, hay escasas posibilidades de construir grupos estables de cooperadores racionales. (Recordemos: los cooperadores de la línea somática que componen la mayor parte de nuestro cuerpo son sistemas intencionales *balísticos,* fiablemente robóticos e insensibles a la tentación, pero ahora estamos hablando de construir no un cuerpo, sino una corporación de individuos altamente racionales, como la Orquesta Sinfónica de Boston.) Y una señal de fiabilidad que sea digna de confianza debe ser, tal como ha demostrado Amotz Zahavi (1987), una señal cara (algo que no sea fácil de simular). Jurar sobre la Biblia es una ceremonia vacía que *no puede* transmitir ninguna información útil, porque si fuera adoptada como señal de fiabilidad, sería inmediatamente copiada y usada por todos los traidores, y, por lo tanto, perdería toda su credibilidad y caería en el desuso. Podríamos tratar de salvarla incrementando la ceremonia —jurar sobre *dos* Biblias, jurar sobre una *pila* de Biblias—, pero la inutilidad de este incremento queda claramente recogida en el refrán,* nuestro paradigma mítico de un intento fallido de demostrar que uno es digno de confianza.[2] Luego el problema principal es: no sólo cómo convertirme en un agente en quien se pueda confiar ante los problemas de compromiso, sino cómo publicitar de manera creíble el hecho de que soy digno de confianza.

En ocasiones un problema se resuelve con otro problema. Esto es especialmente cierto cuando quien se enfrenta al problema es la Madre Naturaleza, la oportunista por excelencia. Nos enfrentamos a un proble-

*** Probablemente se refiera a** *Evett the devil will swear on a stack ofBibles* **[«Incluso el diablo querrá jurar sobre una pila de Biblias»]. (N.** *del t.)*

2. Entonces, ¿por qué persiste la práctica de prestar juramento sobre la Biblia? Porque, con independencia —al menos *hoy***— de la creencia del participante en el castigo divino, indica que uno se pone deliberadamente en posición de cometer perjurio, y asume el riesgo variable pero aún sustancial de un castigo mundano.**

ma de autocontrol realmente difícil —es decir, caro— de resolver. Según Frank, el hecho de que sea caro de resolver es más una bendición que una desgracia. El episodio de Ulises y las sirenas ejemplifica un problema parecido, e igual que allí el truco consiste en idear alguna forma de atarnos al mástil y taparles las orejas a los marineros con cera para que no podamos actuar siguiendo nuestra inclinación irresistible del momento. (El truco es arreglarlo todo para que «en el tiempo *t*» nuestra voluntad sea *inefectiva*.) Ulises conoce perfectamente los beneficios a largo plazo de adoptar la estrategia de evitar a las sirenas cuando cantan su seductora canción, pero también conoce su disposición a sobrevalorar los beneficios inmediatos en numerosas circunstancias, de modo que necesita protegerse de una escala de preferencias en cierto modo viciada que según sus previsiones se le impondrá cuando llegue el momento *t*. Ulises se conoce a sí mismo, y sabe cuál es el recurso que le ha proporcionado la evolución: una facultad más o menos limitada de razonamiento que le hará aceptar el beneficio inmediato («no puedo hacer otra cosa», dirá mientras se lanza a los brazos de las sirenas), a menos que tome ahora las medidas necesarias para distribuir la decisión entre momentos y actitudes más favorables. La seducción de las sirenas no es *inevitable*, si Ulises dispone del tiempo suficiente para preparar una maniobra de evitación. Tal como observa Frank:

> **Es importante subrayar que la literatura experimental no dice que los beneficios inmediatos adquieran un peso"** *excesivo* **en todas las situaciones. Únicamente dice que siempre adquieren un** *gran* **peso. En conjunto, eso era probablemente algo bueno en los entornos donde evolucionamos. Cuando las presiones selectivas son intensas, los beneficios actuales son a menudo los únicos que importan. El presente es, después de todo, la puerta del futuro (Frank, 1988, pág. 89).**

El problema de Ulises no es un problema moral; es un problema prudencial, que también se les plantea a agentes más egoístas y menos altruistas. Para el agente egoísta, el problema consiste en cómo evitar ceder a beneficios egoístas a corto plazo a costa de perder beneficios egoístas a más largo plazo, un problema de lograr un mayor control sobre sí mismo para aumentar su éxito prudencial. Frank sostiene que la resolución de este problema prudencial nos lleva directamente al terreno de la moral, pero antes de abordar su tesis debemos examinar con más detalle el problema de la tentación.

APRENDER A NEGOCIAR CON UNO MISMO

> *La negociación intertemporal parece ser un proceso relativamente artificial que es improbable que surgiera en animales inferiores. Sólo la especie humana encontró la manera de ampliar enormemente el marco de la decisión y descubrió que el libre albedrío nos presta a menudo un peor servicio que la necesidad desnuda.*
>
> GEORGE AINSLIE, *Breakdown ofWill*

Cuando el viejo granjero de Maine comenzó a subirse el mono después de usar el retrete, se le cayó del bolsillo una moneda de 25 centavos que se le fue por el agujero. «¡Diantre!», exclamó, y luego se sacó un billete de 5 dólares de la billetera y lo tiró por el mismo agujero por donde había caído la moneda. «¿Por qué hiciste una cosa así?», le preguntaron. «No pensaréis que iba a rebuscar por allí sólo por una moneda de 25, ¿verdad?», respondió. Aumentar lo que hay en juego modifica el esfuerzo de autocontrol que debemos realizar. Todos solemos tener problemas con la tentación que quedan perfectamente radiografiados con unas sencillas preguntas:

1. *¿Quéprefiriría usted: 1 dólar ahora mismo o 1 dólar mañana?* Si es usted como la mayoría, lo prefiere ahora, por razones evidentes. Cuanto antes lo consiga, antes podrá darle algún uso, y ¿quién sabe lo que le depara el futuro? Si, extrañamente, le diera a usted lo mismo elegir 1 dólar ahora o mañana, la próxima semana o el próximo año, diríamos que usted no *descuenta el futuro.* Obviamente, lo racional es descontar el futuro, pero ¿hasta qué punto?
2. *¿Quéprefiriría usted: 1 dólar ahora mismo o 1,5 dólares mañana?* Si usted prefiere 1,5 dólares mañana, ¿qué diría ante 1,25 dólares? ¿Y 1 dólar y 10 centavos? En algún momento daremos con una elección que le sea indiferente y eso fijará dos puntos en una curva, la curva de descuento del futuro. Podríamos reunir gran cantidad de datos de este tipo para definir múltiples puntos en su curva particular y emplear el dinero como sistema de medida práctico (en sustitución de un conjunto mucho más amplio de preferencias: ¿qué preferiría, no sufrir dolor hoy o no sufrir dolor dentro de una semana? ¿Qué preferiría, ser famoso hoy o ser famoso el año que viene?). Supongamos que usted es indiferente a la pregunta 2. Le parece igualmente deseable 1 dólar hoy o 1,5 dólares mañana. Considere entonces la siguiente pregunta.

3. *¿Qué preferiría usted: 1 dólar el próximo martes o 1,5 dólares el próximo miércoles?* Esta es la misma pregunta que la anterior, sólo que vista desde más lejos en el tiempo. Pero tal vez descubra que sus respuestas no coinciden. Si es usted como la mayoría de la gente, le costará bastante rechazar 1 dólar ahora en favor de *1,5* dólares mañana, mientras que le resultará relativamente fácil ser prudente y firmar pollos *1,5* dólares del miércoles en lugar de hacerlo por el dólar del martes. Si usted tiende a preferir 1 dólar hoy al *,5* dólares mañana, pero al mismo tiempo prefiere 1,5 dólares el miércoles a 1 dólar el martes, tiene usted un conflicto; experimentará usted un cambio en sus preferencias en algún momento del tiempo entre hoy y el próximo martes, un cambio propiciado meramente por el paso del tiempo.

Nuestra susceptibilidad a caer en estos conflictos intertemporales es una pequeña tara, una manía, una anomalía en nuestra competencia básica para la toma de decisiones, y se halla en la base de una notable teoría de la voluntad humana desarrollada por el psiquiatra George Ainslie, a la que recientemente ha dado una formulación accesible en su libro *Breakdown of Will* (2001). La tasa de descuento del futuro puede variar según las personas, y no hay respuesta correcta a la pregunta de cuál debería ser su curva de descuento del futuro, pero sea cual sea ésta en su caso, si usted la aplicara de una manera racional no surgirían conflictos intertemporales: la fría decisión que toma hoy para dentro de un año sería la misma que tomaría una vez pasado el año. Su tendencia a *sucumbir a la tentación* se desvía de su estrategia racional (sea la que sea) de un modo que usted querría evitar racionalmente, si pudiera hacerlo. ¿Qué forma tomaría su curva de descuento? La figura 7.1

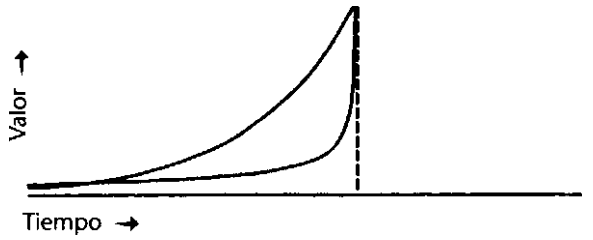

FIGURA 7.1. Una curva de descuento exponencial y una curva hiperbólica (más arqueada) partiendo de la misma recompensa. A medida que pasa el tiempo (en dirección hacia la derecha a lo largo del eje horizontal), el impacto motivacional —*el valor*— de los objetivos del sujeto se acerca a su magnitud no descontada, que queda recogida por la línea vertical (Ainslie, 2001, pág. 31).

muestra dos tipos básicos de curva superpuestos: la curva *exponencial,* más gradual, y la curva *hiperbólica,* muy arqueada y de acusada pendiente.

Puede demostrarse (véase la figura 7.2) que una tasa de descuento exponencial no puede producir dichas anomalías, mientras que una curva de descuento hiperbólica (véase la figura 7.3), al tener una pendiente acusada al final, sí puede producirlas.

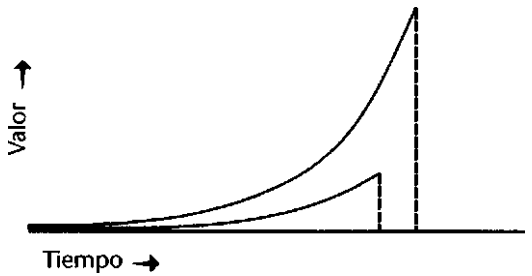

FIGURA **7.2. Curvas de descuento convencionales (exponenciales) desde dos recompensas de distinta magnitud, disponibles en momentos distintos. En todos los puntos donde pueda evaluar el sujeto las recompensas anteriores y posteriores, sus valores se mantendrán proporcionales a sus magnitudes objetivas (Ainslie, 2001, pág. 32).**

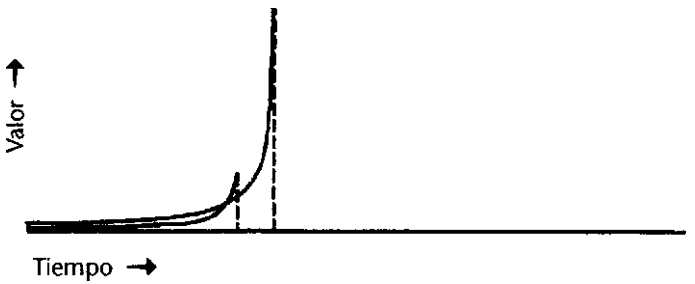

FIGURA **7.3. Curvas hiperbólicas de descuento desde dos recompensas de diferente magnitud disponibles en momentos diferentes. La recompensa menor es preferida durante un breve período de tiempo antes de estar disponible, tal como muestra la porción de su curva que se proyecta por encima de la curva de la recompensa mayor, posterior en el tiempo (Ainslie, 2001, pág. 32).**

En el corto tramo donde el diente de la curva hiperbólica de la recompensa menor se cruza con la curva de la recompensa mayor se abre la ventana de la tentación: un breve período de tiempo durante el cual la recompensa menor parece más valiosa que la mayor. Un número ingente de

pruebas realizadas bajo toda clase de condiciones han demostrado que nosotros, al igual que otros animales, obedecemos de manera innata a ciertas curvas hiperbólicas de descuento. «La especie humana desarrolló evolutivamente una curva de descuento muy regular pero muy arqueada para evaluar el futuro» (Ainslie, 2001, pág. 46). Según Ainslie, se trata de una ilusión muy parecida a la ilusión Müller-Lyer:

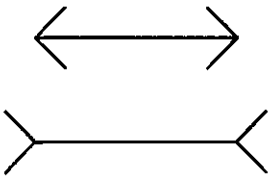

FIGURA 7.4. Ilusión Müller-Lyer.

Es posible que sepamos —por haberlo medido— que las dos líneas tienen la misma longitud, pero eso no impide que la ilusión ejerza toda su fuerza sobre nosotros. Podemos aprender a compensar la ilusión natural mediante una corrección deliberada y consciente. De modo parecido, la teoría de la utilidad (y la medición) puede convencernos de que la tasa de descuento correcta es la exponencial, y podemos aprender a compensar las tasas de descuento hiperbólicas con las que nacimos. Es un acto antinatural, pero que vale la pena aprender a realizar. Algunos lo consiguen más que otros.

Todos nosotros apreciamos al menos vagamente las virtudes de racionalizar nuestro comportamiento de acuerdo con una curva exponencial, pero ¿cómo podemos conseguirlo? ¿De dónde sacaremos la fuerza para dominar nuestros propios instintos? La tradición diría que procede de cierta potencia física llamada ***fuerza de voluntad***, pero esto no hace más que dar un nombre al fenómeno y posponer la explicación. ¿Cómo se implementa la «fuerza de voluntad» en nuestros cerebros? Según Ainslie, la obtenemos a partir de una situación competitiva en la que los «intereses» entran en lo que llama una «negociación intertemporal». Dichos «intereses» vienen a ser una especie de agentes temporales, unos homúnculos que representan varias recompensas posibles:

> Un agente que descuenta hiperbólicamente las recompensas no es un estimador del valor tan simple como se supone que lo es un agente que las descuenta exponencialmente. Viene a ser más bien como una sucesión de estimadores que difieren en sus conclusiones: con el transcurso del tiempo las relaciones recíprocas entre dichos estimadores pasan de la cooperación hacia

un objetivo común a la persecución de fines mutuamente excluyentes. El Ulises que planea el encuentro con las sirenas debe tratar al Ulises que las oye como una persona separada, sobre la que debe influir en lo posible y, en caso contrario, inutilizarla (Ainslie, 2001, pág. 40).

La «negociación de poder» entre esos «grupos de procesos de búsqueda de recompensas» es un proceso que genera su propio equilibrio y que no necesita «ningún ego, juez u otro rey filósofo, ningún órgano de la unidad o la continuidad, aunque sí anticipa cuál sería el aparente funcionamiento de un órgano de este tipo» (pág. 62). Según la descripción que hace Ainslie de este fenómeno, se trata de una competición selectiva en la que los competidores pueden absorberse y explotarse unos a otros, y no es otra cosa (sospecho) que el proceso contradictorio de la «voluntad de perseverar» esbozado por Kane. Sin duda contribuye en gran medida a la impredecibilidad de la facultad de decisión humana, no mediante el uso del azar cuántico, según esperaba Kane, sino mediante la introducción de una recursividad que frustra sistemáticamente las predicciones: cuando decidimos, *usamos reflexivamente nuestra decisión como predicción de cuáles serán nuestras decisiones en el futuro-,* nuestra propia autoconciencia respecto a nuestras decisiones genera un círculo recurrente que hace que nuestras decisiones sean infinitamente sensibles a consideraciones ulteriores.

El ordenado mercado interno figurado por las teorías convencionales de la utilidad se convierte en una confusa pelea sin reglas, donde para prevalecer una opción no sólo debe prometer más que sus competidoras, sino que también debe actuar estratégicamente para impedir que éstas no den la vuelta a la tortilla más adelante (Ainslie, 2001, pág. 40).

Ainslie analiza cómo las microestrategias de esos homúnculos agrupan las recompensas para crear algo parecido a una tasa de descuento exponencial, al generar «reglas» y resoluciones que, a su vez, generan justificaciones para exenciones menores (me resultará más fácil mantener mi dieta si no soy demasiado estricto conmigo mismo, de modo que, como es mi cumpleaños, me recompensaré con un poco de pastel...), que, a su vez, generan medidas y contramedidas ulteriores, un caos de retos internos concatenados. Por ejemplo: «En la medida en que espero conceder excepciones siempre que el impulso sea lo bastante fuerte, ya no tengo una perspectiva creíble de que toda la serie de recompensas posteriores —los beneficios acumulados de mi dieta— estén al alcance de mi decisión. De este modo las curvas de descuento hiperbólicas convierten el autocontrol en una cuestión de autopredicción» (pág. 87).

Supongamos que un alcohólico sometido a una cura de desintoxicación tiene la expectativa de resistirse a la bebida, pero que, sorprendentemente, ve frustrada esta expectativa y, cuando se da cuenta de ello, pierde la confianza en dicha expectativa; si ésta cae tan bajo que deja de ser un contrapeso suficiente para sus ganas de beber, su frustración podría convertirse en una profecía autorrealizadora. Pero si esta posibilidad se ha vuelto lo bastante amenazadora en el período previo a convertirse en la preferida, el sujeto buscará otros incentivos que contraponer a sus ganas de beber antes de que se vuelvan demasiado fuertes, y con ello aumentar su expectativa de no beber, y así sucesivamente (antes de encontrarse tomando un trago). Su decisión está determinada por adelantado, del mismo modo que los eventos tienen causas que a su vez tienen otras causas; pero lo que determina de manera inmediata su decisión es la interacción de elementos que, por más que sean en sí mismos conocidos, vuelven impredecible el resultado cuando interactúan de manera recursiva.

El descuento hiperbólico convierte la toma de decisiones en un fenómeno colectivo, donde el colectivo está integrado por sucesivas disposiciones del individuo a lo largo del tiempo. En cada momento toma la decisión que le parece mejor; pero el cuadro que se le presenta viene representado en buena medida por sus expectativas acerca de qué decisiones tomará en ocasiones posteriores, una expectativa que se funda principalmente en las decisiones que ha tomado en ocasiones previas (Ainslie, 2001, pág. 131).

La teoría de la voluntad de Ainslie genera explicaciones para un buen número de fenómenos que han dejado perplejos a otros teóricos (o que han sido simplemente ignorados por ellos), sobre temas tales como la adicción y la compulsión, el «saciado prematuro», el autoengaño y la desesperación, el pensamiento «legalista» y la espontaneidad. El precio que se debe pagar por esta fecundidad teórica son unas premisas inicialmente contraintuitivas: en particular, la necesidad de distinguir entre placeres y recompensas. Una recompensa es, por definición, «cualquier experiencia que tienda a incitar la repetición del comportamiento que la precede», a pesar de que algunas de dichas experiencias son positivamente dolorosas, por mucho que promuevan la disposición a la replicación (por mucha que sea la aptitud intracerebral, podría decirse) de dicho comportamiento. Es una teoría difícil, plagada de novedades que exigen abandonar viejos hábitos de pensamiento, y en esta presentación no he hecho sino esbozar sus conclusiones más interesantes. Es una teoría que no ha recibido aún la atención que merece, por lo que sigue abierta la cuestión de cuáles de sus muchas conclusiones tentadoras merecen nuestra sanción, pero no hay

duda de que es una magnífica aportación a los recientes trabajos que han venido a aplicar la perspectiva evolucionista a las cuestiones filosóficas tradicionales relativas a la voluntad y a la mente. De la teoría se desprenden incluso algunas observaciones inquietantes sobre el tema del carácter elusivo de la moral y sobre cómo nuestras reglas mejor formuladas pueden volverse en nuestra contra y dar lugar a consecuencias no deseadas, aunque éstos son temas que deberán quedar para otra ocasión. Todavía no hemos llegado al terreno de la moral, aunque Robert Frank propone un camino para lograrlo.

NUESTROS CAROS EMBLEMAS AL MÉRITO

Supongamos que ponemos un caramelo delante de un niño pequeño y le decimos que puede comérselo, pero que si es capaz de esperar quince minutos, podrá comerse dos. ¿Cómo responden los niños al desafío de posponer la gratificación? No demasiado bien. Los niños manifiestan importantes variaciones en su capacidad de autocontrol, las cuales sin embargo no son *inevitables*, ya se deban a diferencias genéticas, a diferencias en el entorno durante la primera infancia, o al mero azar; pueden recortarse (o aumentarse) mediante unas sencillas estrategias de distracción o concentración. (Por ejemplo, los niños pueden aprender a esperar al segundo caramelo si se concentran en las deliciosas propiedades de algo que no tienen a su alcance: unas crujientes rosquillas saladas, por ejemplo, o su juguete favorito.) Algunas estrategias apelan a la fría razón, y otras a la competencia de la pasión. Estas propuestas de manipulación de uno mismo se enfrentan, por cierto, a una influyente doctrina en filosofía moral, atribuida a Kant, que subraya el carácter bajo e innoble de dichas muletas *meramente emocionales*. El ideal kantiano es una fantasía según la cual deberíamos ser capaces de reforzar nuestro músculo del razonamiento puro hasta el punto de poder realizar juicios puros y vacíos de emoción, sin mácula de chabacanos sentimientos de culpabilidad o deseos básicos de amor o aceptación. Kant sostenía que dichos juicios no sólo son la mejor versión de nuestros juicios morales, sino que constituyen la única clase de juicios que pueden considerarse propiamente morales. Estimular la reflexión con remisiones a emociones básicas puede ser una buena manera de formar a los niños, pero la presencia de dichas estrategias en su formación descalifica sus juicios de cualquier posible consideración moral. ¿Se trata tal vez de un ejemplo de cómo la búsqueda de la perfec-

ción —una deformación profesional en el caso de los filósofos— impide ver el mejor camino?

Según Frank, la belleza evolutiva de este aprovechamiento de la emoción para el autocontrol es que al mismo tiempo sirve para ofrecer la cara señal que se requiere para dar a conocer este mismo triunfo: los demás se dan cuenta de que somos uno de esos tipos emotivos en quienes se puede confiar, porque se toman sus compromisos con *pasión-,* no es que estemos *locos* o seamos *irracionales,* sino que atribuimos un precio irracionalmente alto (desde la miope perspectiva del crítico) a nuestra integridad. Llevamos el corazón prendido de la manga como un emblema, y bien caro que nos cuesta. El truco para ganarse la reputación de ser bueno, un premio muy valioso, es ser realmente bueno. Ningún método más barato funcionará mejor (aunque... la evolución sigue adelante).

Para comprender por qué *ser realmente bueno* es la solución más económicamente eficiente para este problema, debemos verlo como el precio que pagamos por el autocontrol. Sólo puedo controlarme a mí mismo usando una brocha gorda. «Los sentimientos morales pueden verse como un burdo intento de afinar el mecanismo de las recompensas, para hacerlo más sensible a recompensas distantes y a ciertas penalizaciones en circunstancias señaladas» (Frank, 1988, pág. 90). Tal como veremos en el próximo capítulo, no puedo controlar al detalle y en cada momento todas mis deliberaciones, de modo que debo recurrir a fórmulas más expeditivas y equiparme con poderosas disposiciones emocionales que desbordan sus objetivos y me dejan temblando de rabia cuando la rabia es lo que corresponde, incapaz de contener mi alegría cuando la alegría es lo que corresponde, o bien hundido en la pena o en la compasión. Pero para conseguir que estas emociones me ayuden a tomar decisiones prudenciales a largo plazo cuando me enfrento a la tentación a corto plazo de las sirenas, debo permitirles que me dominen también cuando mi elección es entre una ganancia a corto plazo y un beneficio para los demás. No puedo comprometerme *únicamente* conmigo mismo. O, para expresarlo según mi lema, el entorno social en el que me encuentro me anima a hacerme más grande de lo que sería, precisamente para fomentar mis *estrechos* intereses personales; cuando «trato de ser el número uno» despliego una red lo bastante grande como para incluir a aquellos que cooperan conmigo.

Como siempre, no es suficiente postular este feliz estado de cosas como si fuera un regalo de Dios. Tal vez se dé a veces por accidente, pero sí persiste lo bastante como para convertirse en una pauta dentro del mundo, requiere una explicación. La tarea de los modelos evolutivos es

demostrar que existe la posibilidad de que evolucionen entornos en los que esta clase de extensión del yo sea una maniobra necesaria, dictada por la razón. Esta «decisión» en el terreno del diseño —pagar el precio de comprometerse con una serie de muestras de altruismo *impuro* (¿o es sólo benegoísmo avanzado?) a cambio de alcanzar un mayor autocontrol— viene avalada por unas razones que nadie tiene por qué haber apreciado. Son razones virtuales, pero no por ello dejan de serlo. De hecho, es *mejor* que sean virtuales. Eso es lo que confiere a la expresión emocional su valor como evidencia en la carrera de armamentos del engaño y la detección. Si como individuos pudiéramos identificar esas acciones y actuar de acuerdo con ellas, concentrar en ellas nuestras mentes, los demás sospecharían que estamos haciendo teatro. Somos unos jueces del carácter muy atentos, y un examen de los indicios que más nos importan (nos demos cuenta o no conscientemente de ello) revelaría que prestamos escasa atención a las demostraciones que resultan fáciles de simular y que, en cambio, nos concentramos en las señales que constituyen manifestaciones irreprimibles e inimitables de una disposición. Y eso es lo único que vemos, según Frank:

> **Podemos imaginar, por lo tanto, una población donde las personas con conciencia tienen más éxito que aquellas que no la tienen. La gente que carece de ella haría trampas menos veces si pudiera, pero simplemente tiene mayores dificultades para resolver el problema del autocontrol. La gente que tiene conciencia, en cambio, es capaz de adquirir una buena reputación y cooperar con éxito con otros individuos de disposición parecida (Frank, 1988, págs. 82-83).**

¿Dónde deja todo esto el contraste entre el benegoísmo y el genuino altruismo? Frank pretende que la innovación por él descrita llega hasta la línea de meta y nos permite alcanzar el genuino altruismo:

> **Las personas con genuinos sentimientos morales son más capaces que las otras de actuar según su propio interés [...]. Las personas con buena reputación pueden resolver de este modo incluso dilemas del prisionero no reiterados. Por ejemplo, pueden cooperar con éxito unas con otras en situaciones en las que el engaño sería imposible de detectar. En otras palabras, el genuino altruismo puede surgir meramente sobre la base de haber establecido una reputación de comportarse de manera prudente (pág. 91).**

Frank pone de relieve que los altruistas —si es que esa buena gente es realmente altruista— llegan a tener bastante éxito, a pesar de los costes en

los que incurren. Psicólogos y economistas han realizado numerosos experimentos que enfrentan a seres humanos (normalmente sus compañeros de estudios) a múltiples dilemas del prisionero en los que las retribuciones son sumas de dinero pequeñas, aunque no negligibles. En los experimentos que realizó Frank, se ofrecía a los estudiantes diversas oportunidades de conocerse unos a otros en el curso de breves encuentros (entre diez minutos y media hora) antes de emparejarlos repetidas veces en interacciones del tipo del dilema del prisionero. Mediante la introducción de variaciones en las condiciones, Frank demostró que la gente es sorprendentemente buena —aunque dista de ser perfecta: entre un 60 y un 75 % de acierto— a la hora de predecir quién traicionará y quién cooperará.

> El experimento del dilema del prisionero apoya nuestra intuición de que podemos identificar a las personas no oportunistas. Que seamos capaces de hacerlo es en realidad la premisa en la que se basa todo nuestro modelo de compromiso. De esta premisa se sigue lógicamente que el comportamiento no oportunista podrá surgir y sobrevivir incluso en un mundo competitivo, materialista y cruel. Podemos conceder, pues, que las fuerzas materiales son las que gobiernan en último término el comportamiento, pero al mismo tiempo rechazar la idea de que la gente esté movida siempre y en todas partes por el interés material (Frank, 1988, pág. 145).

> Tal como subrayan los racionalistas, vivimos en un mundo material donde, a la larga, termina por dominar el comportamiento más conducente al éxito material. Una y otra vez, sin embargo, vemos que los comportamientos con mayor éxito selectivo no surgen directamente de la persecución de ventajas materiales. A causa de importantes problemas de compromiso e implementación, dicha persecución demuestra a menudo ser contraproducente. Para obtener buenos resultados, a veces debemos dejar de preocuparnos por sacar el máximo beneficio (pág. 211).

Varios aspectos de la teoría de Frank sugieren sorprendentes correcciones al viento filosófico dominante que hemos encontrado en capítulos precedentes. En primer lugar, recordemos la discusión del capítulo 4 acerca de si «podríamos haber hecho otra cosa», y el ejemplo de Martin Luther. Lejos de considerar que dichos fenómenos sean excepciones a la regla, o casos especiales necesitados de excusas especiales, podemos ver que la práctica de ponerse a uno mismo en una posición en la que no podía hacer otra cosa es una innovación clave en el avance evolutivo por el Espacio del Diseño —el espacio Vasto y multidimensional de todos los di-

seños posibles— hacia la libertad humana. Podemos descubrir un rastro fósil de esta táctica de fijar la propia voluntad, una vez identificada, en una palabra de elogio moral que raramente recibe la atención de los filósofos pero que a menudo es motivo de consideración en un agente moral: demuestra mucha *determinación*, decimos de alguien con admiración. En segundo lugar, el miedo de los filósofos a que si estamos determinados tal vez no seamos capaces de aprovechar las verdaderas oportunidades —o a que si estamos determinados tal vez no *haya* verdaderas oportunidades— deja paso, como hemos visto, a la posibilidad contraria: sólo podemos ser libres en un sentido moralmente relevante si aprendemos a hacernos *insensibles* a muchas de las oportunidades que se nos presentan. De nuevo, no hacemos esto a base de volvernos locos o ciegos, sino aumentando nuestra apuesta, para que las «decisiones» se conviertan en movimientos forzados, en obviedades que no precisan consideración seria. En tercer lugar, hemos visto que aquel ser mitológico, el agente racional puramente egoísta de los economistas que nunca es capaz de resistirse a una ganga, es un tonto racional a quien se le podría hacer la famosa pregunta retórica: «¿Cómo puede ser usted tan rico, siendo tan estúpido?». En palabras de Frank:

> **Los altruistas [...] parecen tener mayor éxito económico: los estudios experimentales revelan de manera consistente que el comportamiento altruista se ve positivamente correlacionado con el estatus socioeconómico. Por supuesto, esto no significa que el comportamiento altruista lleve consigo necesariamente el éxito económico. Pero sí sugiere que una actitud altruista no resulta demasiado perjudicial en términos materiales (Frank, 1988, pág. 235).**

A otro ser mitológico, el santo racional kantiano, le podemos responder en la misma línea: «Si somos tan inmorales, ¿cómo es que tenemos tantos amigos que confían en nosotros?». En otras palabras, si queremos llegar hasta el genuino altruismo, lo mejor es adoptar la vía evolucionista e ir ascendiendo por ella mediante incrementos graduales, sin Mamíferos Primordiales ni ganchos colgados del cielo, y pasar del egoísmo ciego al pseudoaltruismo y luego al cuasi altruismo (benegoísmo), para llegar finalmente a algo con lo que tal vez nos demos por satisfechos la mayoría de nosotros.

Permítanme hacer una breve reflexión sobre los métodos que he recomendado y las conclusiones a las que *no* he llegado. Los argumentos y

las conclusiones de Frank no han logrado ni mucho menos una aceptación general entre sus colegas economistas o teóricos de la evolución (o filósofos), y siguen en pie importantes problemas —y alternativas— que deberán ser cuidadosamente examinados. Lo que me parece más importante en este punto es que el proyecto de Frank, igual que el de Ainslie, es un ejemplo de un *tipo* de planteamiento de raíz darwinista que resulta en mi opinión tan preceptivo como prometedor para tratar estas cuestiones. Es preceptivo porque cualquier teoría ética que se limite a establecer cómodamente una lista de virtudes humanas sin tratar de explicar cómo pueden haber surgido corre el peligro de presuponer algún gancho colgado del cielo, algún milagro que no «explica» nada precisamente porque puede «explicarlo» todo. Resulta prometedor porque, al revés de lo que declaran los enemigos del darwinismo, las propuestas más recientes echan por el suelo muchas veces las doctrinas de dichos teóricos. Los ejercicios especulativos sobre el diseño de los agentes han sido el pan de cada día de los filósofos desde *La república* de Platón. Lo que aporta la perspectiva evolucionista es una forma relativamente sistemática de mantener estos ejercicios dentro del terreno naturalista (de modo que no terminemos diseñando un ángel o un móvil perpetuo), y lo que es igual de importante, la posibilidad de explorar una serie de interacciones entre agentes a lo largo del tiempo frente a las cuales los filósofos acostumbran a escurrir el bulto. Por ejemplo, los filósofos acostumbran a preguntar, retóricamente: «¿Qué pasaría si todo el mundo lo hiciera?», sin pararse a considerar la respuesta, que por lo general estiman evidente. Nunca se les ocurre plantear otra pregunta mucho más interesante: ¿qué ocurriría si *algunas* personas lo hicieran? (¿Qué porcentaje, a lo largo de qué período de tiempo, y bajo qué condiciones?) Las simulaciones informáticas de escenarios evolutivos aportan una nueva disciplina: una forma de descubrir las premisas ocultas de los propios modelos, y una forma de explorar los efectos dinámicos, de «jugar con los botones» para ver el efecto que tienen los cambios de entorno sobre las variables. Es importante darse cuenta de que dichas simulaciones informáticas son en realidad experimentos mentales filosóficos, formas de generar ideas, no experimentos empíricos. Exploran sistemáticamente las implicaciones de ciertos conjuntos de premisas. Hasta ahora los filósofos debían realizar los experimentos mentales a mano, uno por uno. Ahora pueden introducir miles de variaciones en una hora, una buena forma de comprobar si sus intuiciones son el resultado de algún elemento arbitrario del entorno.

Hemos conseguido esbozar —sólo esbozar— un camino que lleva desde el origen de la vida hasta el surgimiento de personas, de unos agentes cuya libertad es a un tiempo su mayor fuerza y su principal problema. Ahora debemos examinar más de cerca lo que debe ocurrir en el interior de tal agente humano cuando toma una decisión libre, antes de examinar las implicaciones de la evolución aún en curso de la libertad humana.

Capítulo **7**

La complejidad de la vida social en una especie poseedora de lenguaje y cultura da lugar a una sucesión de carreras armamentísticas evolutivas como resultado de las cuales los agentes desarrollan algunos componentes clave de la moral humana: un interés por descubrir condiciones que promuevan la cooperación, una sensibilidad hacia los castigos y las amenazas, una preocupación por la propia reputación, unas marcadas disposiciones hacia la automanipulación, diseñadas para mejorar el autocontrol frente a la tentación, y la capacidad de adoptar compromisos apreciables para los demás. Innovaciones como éstas pueden tener éxito bajo condiciones especificables que coevolucionan con ellas y terminan por sustituir el «egoísmo» miope de organismos más simples que habitan nichos más simples.

Capítulo 8

La imagen emergente del agente humano como un enjambre de intereses en competencia diseñados por fuerzas evolutivas resulta difícil de conciliar con nuestra percepción tradicional de nosotros mismos como egos, almas o yoes conscientes, que determinamos nuestras acciones intencionales mediante decisiones libres que deben proceder de nuestros santuarios privados de la mente. Esta tensión queda claramente reflejada en un controvertido —y a menudo malentendido— experimento de Benjamín Libet, y puede resolverse si examinamos con más detalle la emergencia del yo a partir de los procesos que tienen lugar en nuestros cerebros. Corregir estos malentendidos tan comunes sobre el yo y el cerebro hace que se desvanezcan también algunas sombrías conclusiones acerca del futuro de la libertad que han ganado crédito en algunos círculos.

NOTAS SOBRE FUENTES Y LECTURAS COMPLEMENTARIAS

Entre los muchos libros excelentes dedicados a las teorías evolucionistas de la cooperación cabe citar *Evolution of the Social Contract* (1996), de Brian Skyrms; *The Moral Animal* (1994) y *Nonzero* (2000), de Robert Wright; *The Origins of Virtue* (1996), de Matt Ridley; *Sex and Death: An Introduction to Philosophy of Biology* (1999), de Kim Sterelny y Paul E. Griffiths; y, por supuesto, *Unto Others* (1998), de Elliott Sober y David Sloan Wilson. Para un valioso comentario del libro de Sober y Wilson (y una réplica), véase Katz, 2000. He expresado mis opiniones acerca de su libro en un artículo que aparecerá próximamente en *Philosophy and Phenomenological Research* (Dennett, en preparación a) y que también contendrá algunos comentarios más y una réplica de los autores.

Sobre el sencillo modelo de castigo necesario para hacer respetar las normas culturales, véase *Having Thought* (1999), de John Haugeland, y mi reseña (Dennett, 1999a). Paul Bingham (1999) ha desarrollado una atrevida y controvertida teoría de la evolución humana basada en la premisa de que la invención de armas sencillas —palos y piedras— alteró hasta tal punto las relaciones de coste-beneficio o el riesgo de la participación individual en los castigos del grupo contra los traidores que propició las variedades únicas de cooperación social humana de las que depende la cultura humana, una revolución cultural evolutiva que rápidamente tuvo respuesta a nivel genético, a través de adaptaciones del esqueleto para mejorar su capacidad de arrojar piedras y blandir armas.

El Principio de Zahavi aparece discutido por extenso en Frank, 1988. Véase también *The Ant and the Peacock* (1991), de Helena Cronin. Randolph Nesse ha editado una imponente antología de trabajos recientes sobre el tema del compromiso, en *Evolution and the Capacity for Commitment* (2001).

Para un repaso general a la literatura experimental sobre el tema de la automanipulación y el autocontrol entre los niños, véase «A Hot/Cool System Analysis of Delay of Gratification: Dynamics of Willpower» (1999), de J. Metcalfe y W. Mischel. Para un repaso general al contexto de teoría de juegos donde se enmarca la propuesta de Frank, así como una crítica sutil y algunas correcciones amistosas a su invocación de las emociones para cumplir la función de señales, véase «Emotions as Strategic Signals», de Don Ross y Paul Dumouchel.

Capítulo 8

¿Está usted fuera de la cadena?

> *Imaginas un constructo mental ficticio llamado «libre albedrío», lo que para un neurocientifico cognitivo es algo así como creer en elfos o en ovnis.*
>
> RACHEL PALMQUIST, **un personaje de** *Brain Storm,* **de Richard Dooling**

Hace algunos años tuve una extraña experiencia. Estaba leyendo una novela divertida e intelectualmente estimulante de Richard Dooling, titulada *Brain Storm* (1998), que me había recomendado un amigo que insistía en que me gustaría, a pesar del título (en 1978 yo había publicado un libro titulado *Brainstorms).*

SACAR LA MORALEJA EQUIVOCADA

El héroe de esta novela es un joven abogado que visita un laboratorio neurocientifico en el intento de demostrar que su cliente, acusado de asesinato, tiene una lesión cerebral. La neurocientífica que se presta a ayudarle, la doctora Rachel Palmquist, es —¡cómo no!— tan hermosa como desinhibida y las cosas terminan por subir de tono. Desaparecen sus ropas respectivas y, cuando ambos se encuentran ya entrelazados en el suelo del laboratorio, se encuentran con un problema: nuestro héroe, según parece, tiene conciencia, y los pensamientos relativos a su esposa y a sus hijos amenazan con poner un abrupto final al comercio carnal. ¿Qué hacer? La doctora Palmquist hace lo que supongo haría cualquier neurocientífica desnuda y competente bajo las mencionadas circunstancias. Dice:

> **En** *La conciencia explicada***, Dan Dennett usa la analogía del dibujo animado de Casper. Lo que quieres decir es que tienes alma (Dooling, 1998, pág. 228).**

La cuestión objeto de debate es la libertad y, según dice ella, yo explico que no puede existir:

> **—¿Ni siquiera somos libres?**
>
> **—Psicología popular otra vez —dijo ella—. Es una bonita ficción. Tal vez una ficción necesaria: la posibilidad de que una parte de tu conciencia pueda separarse de sí misma y, desde su nuevo emplazamiento, evaluar y controlar su propia actividad. Pero un cerebro es una orquesta sinfónica sin director. En este momento estamos oyendo un oboe o tal vez un flautín que realiza una fioritura de autoexamen, mientras el resto de los instrumentos se lanzan a un *crescendo* totalmente distinto. Lo que queda de ti se encuentra en un equilibrio extremadamente complejo de procesadores paralelos biológicos húmedos en competencia entre sí dentro de esa olla electroquímica de fideos que fermenta entre tus orejas, que está a cargo de tu cuerpo, pero que por definición no puede hacerse cargo de sí misma (Dooling, 1998, pág. 229).**

¡Todo un discurso! Sin duda debe tratarse de una neurocientífica brillante, puesto que a continuación procede a ofrecer un improvisado resumen de mi teoría de la conciencia que resulta sugerente y muy correcto —lo que ya es difícil de conseguir con la ropa puesta y detrás de un podio—, pero lo que más me gustó fue el toque maestro de Dooling: su personaje interpreta mal la parte de la libertad, *exactamente en el mismo sentido en que lo han hecho algunos neurocientíficos de verdad.* La cuestión es: ¿opino realmente que el libre albedrío es una ficción? ¿Es eso lo que se desprende de mi teoría de la conciencia? En absoluto, pero más de unos cuantos neurocientíficos y psicólogos piensan que su ciencia ha demostrado precisamente esto, y mi alusión a Casper puede haber contribuido a este malentendido.

Resulta más fácil comprender lo que quiero decir si cambiamos por un momento de fantasía. Recordemos el mito de Cupido, que revolotea con sus alas de querubín, dedicado a enamorar a las personas con su pequeño arco y sus flechas. Se trata de una convención tan poco convincente que resulta difícil creer que alguien la haya tomado en serio en alguna de sus versiones. Pero podemos pretenderlo: supongamos que hubo alguna vez personas que creían que si la gente se enamoraba era por causa de una flecha invisible lanzada por un dios volador. Y supongamos que algún científico aguafiestas llegó y demostró que simplemente no era verdad: no existía ningún dios volador de este tipo. «Ha demostrado que nadie se enamora jamás, no *de verdad.* La idea de enamorarse no es más que una bonita (tal vez necesaria) ficción. Nunca ocurre.» Eso es lo que dirían

algunos. Cabe esperar que otros lo negarían: «No. El amor es bien real, y también los flechazos amorosos. Simplemente no es lo que la gente pensaba que era. Es igual de bueno, tal vez incluso mejor. El verdadero amor no requiere ningún dios volador». Ocurre algo parecido con la cuestión de la libertad. Si es usted uno de los que piensan que la voluntad sólo es verdaderamente libre si procede de un dios inmaterial que flota felizmente en su cerebro y lanza flechas de decisión sobre su córtex motor, dado el sentido que da usted a la libertad, mi opinión es que no existe. Si, en cambio, usted piensa que podría haber libertad moralmente relevante sin necesidad de que fuera sobrenatural, mi opinión es que la libertad es real, pero no es exactamente lo que habíamos pensado que era.

Como siempre hay lectores situados en ambos bandos, es imposible llegar a todo el mundo si no se dedica un esfuerzo especial a llamar la atención de todos sobre este problema, cosa que he tratado de hacer repetidamente. Una de las cuestiones que trataba en mi libro *Brainstorms* era si las *creencias* o los *dolores* eran «reales», para lo cual inventé una pequeña fábula sobre unas personas que hablaban un idioma en el que decían sufrir «fatigas» cuando nosotros diríamos que estamos cansados o exhaustos. Cuando entramos nosotros en escena con nuestra sofisticada ciencia, dichas personas nos preguntan cuáles de las pequeñas cosas que circulan por su sangre son las fatigas. No le vemos sentido a la pregunta, lo que les lleva a exclamar, incrédulos: «¿Niega usted que las fatigas sean reales?». Dada su tradición, se trata de una pregunta difícil de responder, que requiere diplomacia (no metafísica) por nuestra parte. En *La conciencia explicada* (1991a), traté de deshacer la misma confusión con la historia de un loco que decía que no había animales en el zoo (sabía perfectamente que había jirafas y elefantes y demás, pero insistía en que no eran lo que la gente pensaba que eran). Me parecía que había suficiente con dichos ejercicios de imaginación para resolver la cuestión, pero debo decir que el mensaje no parece haber llegado. Finalmente me he dado cuenta de que a mucha gente le gusta mantener el equívoco. No quieren corregir sus imaginaciones. Les gusta decir que yo niego la existencia de la conciencia, que yo niego la existencia del libre albedrío. Incluso un pensador de la inteligencia de Robert Wright encuentra irresistible la negación de la distinción que propongo:

> **Por supuesto, el problema es aquí la tesis de que la conciencia es «idéntica» a los estados físicos cerebrales. Cuanto más se esfuerzan Dennett y otros por explicarme lo que quieren decir con eso, más me convenzo de que lo que realmente quieren decir es que la conciencia no existe (Wright, 2000, pág. 398).**

Y ese penetrante observador cultural que es Tom Wolfe señala que E. O. Wilson, Richard Dawkins y yo mismo presentamos

> **elegantes argumentos para explicar por qué la neurociencia no reduce en lo más mínimo la riqueza de la vida, la magia del arte, la justicia de las causas políticas [...]. Por muchos esfuerzos que hagan, sin embargo, la neurociencia no es recibida por el público como una gran promesa salida de las universidades. Pero sí llega al público, y rápidamente. La conclusión que saca la gente fuera de las paredes del laboratorio es:** *¡todo está fijado de antemano! ¡Somos todos máquinas! Eso, y: ¡no me echéis la culpa a mí! ¡Tengo mal los circuitos!* **(Wolfe, 2000, pág. 100).**

Exactamente la misma conclusión a la que Rachel Palmquist pretendía llegar en el suelo del laboratorio. Más adelante en este capítulo nos enfrentaremos al problema de cara, tal como queda formulado en el título de un excelente libro del fisiólogo Daniel Wegner: *The Illusion of Conscious Will* (2002). La teoría de Wegner sobre la voluntad consciente es la mejor que conozco, y coincido con ella en casi todos los puntos. He hablado con él sobre su extraña elección del título (desde mi punto de vista). Wegner adopta en mi opinión el papel del científico aguafiestas que demuestra que Cupido no lanza ninguna flecha y luego insiste en titular su libro *La ilusión del amor romántico.* Pero soy consciente de que algunas personas insistirán en que el título de Wegner dice las cosas tal como son: lo que Wegner demuestra es *en efecto* que la voluntad consciente es una ilusión. Wegner suaviza el golpe al final diciendo que, aunque la voluntad consciente sea una ilusión, la acción responsable y moral es real. Y eso es lo esencial para los dos. Estamos de acuerdo en que Rachel Palmquist se equivoca cuando usa una teoría neurocientífica de la voluntad para fundamentar su conclusión de que la conciencia de nuestro héroe no debería causarle ningún problema (puesto que en realidad no es libre). Wegner y yo coincidimos en lo esencial; estamos en desacuerdo en cuanto a la estrategia. Wegner piensa que es menos ambiguo, más efectivo, decir que la voluntad consciente es una ilusión, aunque sea una ilusión benigna, y en ciertos aspectos una ilusión verídica. (¿No es esto una contradicción en términos? No necesariamente; del mismo modo que hablamos de un átomo divisible, a pesar de su etimología, nada impide que una ilusión verídica pueda encontrar un lugar en nuestro esquema conceptual.) Por mi parte pienso que la tentación de interpretar esta conclusión en el sentido de Rachel Palmquist es tan fuerte que prefiero formular las mismas tesis diciendo que no, que la libertad no es una ilusión; todas las formas de libre

albedrío valiosas y deseables están, o pueden estar, a nuestra disposición (pero debemos abandonar cierta ideología falsa y anticuada para comprenderlo). El amor romántico sin la flecha de Cupido sigue siendo algo por lo que vale la pena suspirar. Sigue siendo amor romántico, auténtico amor romántico.

CONRAD: ¡No, no lo es! ¡El amor romántico sin genuina espiritualidad —lo que usted caricaturiza como la flecha de Cupido— no es auténtico amor romántico! ¡Es un sustituto barato! Y lo mismo puede decirse del libre albedrío. Lo que usted llama libre albedrío, un fenómeno que en último término no es sino un complicado engranaje de causas mecánicas que *parece* una decisión libre (vista desde ciertos ángulos), ¡pero que no lo *es* en absoluto!

De acuerdo, Conrad, si insiste en usar los términos de este modo. Pero le corresponde a usted demostrar por qué es mejor mantener estas formas «genuinas» del amor romántico y del libre albedrío, cuando los sustitutos que propongo cumplen con todos los requisitos que ha enumerado usted hasta ahora. ¿Por qué habríamos de preocuparnos por las formas «genuinas»? Estoy de acuerdo en que la margarina no es auténtica mantequilla, por muy bueno que sea su sabor, pero si usted insiste en tener mantequilla auténtica a cualquier precio, debería dar una buena razón para ello.

CONRAD: ¡Ajá! Entonces lo admite. Usted no hace sino jugar con las palabras y tratar de colar la margarina como si fuera auténtica mantequilla. ¡Animo a todo el mundo a exigir auténtica libertad; no acepten sustitutos!

¿Y aconseja usted también a los diabéticos que insistan en tener «auténtica» insulina, y no un sustituto «artificial»? ¿Si su auténtico corazón falla alguna vez, rechazará usted un sustituto artificial capaz de realizar todas las funciones de su corazón real? ¿En qué punto el amor por la tradición se convierte en una estúpida superstición? Sostengo que las modalidades de libre albedrío que estoy defendiendo merecen la pena precisamente porque todas desempeñan las funciones *valiosas* que tradicionalmente se han atribuido al libre albedrío. Pero no puedo negar que la tradición asigna también otras propiedades al libre albedrío que no están presentes en las modalidades que propongo. Y yo digo: tanto peor para la tradición.

El tiempo dirá tal vez qué estrategia expositiva, si la de Wegner o la mía, es la más apropiada para el tema de la libertad, o tal vez no. Pero vergüenza debería darles a todos aquellos que ignoran la tesis —explícitamente defen-

dida por ambos— de que una teoría naturalista de la toma de decisiones sigue dejando mucho margen para la responsabilidad moral.[1]

¿Qué es lo que tiene la explicación neurocientífica de la toma de decisiones que convence a tanta gente de que la libertad es una ilusión? No es el mero hecho del materialismo —el hecho de que no hay Cupidos que disparen flechas sobre nuestro córtex motor—, sino un aspecto particular de la explicación neurocientífica, y Rachel Palmquist expresa magníficamente la impresión popular:

> **La cognición preconsciente es aquella actividad cerebral que tiene lugar *antes* de que seas consciente de ella. Lo que tal vez no te gustará saber es que esta actividad genera movimientos en el mundo físico. Tu conciencia, si quieres llamarla así, se limita a observar una actividad que se origina en otra parte del cerebro [...]. Piensa en tu cerebro como un complejo sistema de redes y procesadores en paralelo. De vez en cuando, algunos son conscientes de ellos mismos, pero la mayoría no lo son. Imagina un vacío moral de 300 milisegundos que se abre justo después de que el cerebro active el comportamiento y antes de que el cerebro sea consciente de ello (Dooling, 1998, pág. 120).**

El problema es ese «vacío moral» de 300 milisegundos. ¡Parece como si el cerebro tomara sus decisiones antes de que lo hiciéramos nosotros!

> **«Estímulos, sensaciones —dijo ella, mientras colocaba un electrodo en cada hombro—. Su procesamiento es preconsciente, por lo que hay importantes decisiones y representaciones mentales que se llevan a cabo antes de que el cerebro sea consciente de ellas» (pág. 122).**

El «agujero» de los 300 milisegundos es sin duda real, pero hay algo sospechoso en esta forma de interpretarlo —como un «agujero moral»—, y éste es el error que me interesa examinar. De nuevo. Ya lo traté en un capítulo de *La conciencia explicada,* pero era una exposición oscura y difícil, y será bueno que la refresque un poco. Tal vez ahora la moraleja de la his-

1. Discrepa de nosotros Derk Pereboom, cuyo último libro, *Living without Free Will* (2001), salió justo cuando yo estaba dando los últimos retoques a este libro. Pereboom defiende la tesis de que «de acuerdo con nuestras mejores teorías científicas, nuestras acciones están causadas en último término por factores que están más allá de nuestro control y, por lo tanto, no somos responsables de ellas». No logró persuadirme en lo más mínimo, pero tal vez aquellos que no estén convencidos con mi libro puedan encontrar un valioso aliado en él.

toria llegue con más claridad, y no al revés, tal como la interpretó Rachel Palmquist, esa neurocientifíca brillante y desnuda.

SIEMPRE QUE EL ESPÍRITU SE LO PIDA

> *¿Son voluntarias las decisiones? ¿O son cosas que nos ocurren? Desde ciertos puntos de vista parecen movimientos preeminentemente voluntarios en nuestras vidas, los instantes en los que ejercemos plenamente nuestra agencia. Pero esas mismas decisiones pueden dar la extraña impresión de estar fuera de nuestro control. Debemos esperar para ver cuál será nuestra decisión sobre algo, y cuando finalmente la tomamos, nuestra decisión emerge a la superficie de nuestra conciencia desde no se sabe dónde. No somos testigos de su proceso de* **formación;** *sólo somos testigos de su* **llegada.** *Esto puede llevar a la extraña idea de que el cuartel general no se encuentra allí donde estamos nosotros, como sujetos conscientes e introspectivos; se halla en algún lugar más profundo, e inaccesible para nosotros.*
>
> DENNETT, *Elbow Room*

El cerebro tarda un tiempo en hacer las cosas, de modo que siempre que *usted* hace algo (siempre que su cuerpo hace algo), su cerebro, que es quien controla su cuerpo, tiene que haber hecho algo primero. Normalmente, cuando estamos despiertos y ocupados, hacemos varias cosas a la vez: caminar y hablar, remover el contenido de la cacerola sobre el fogón mientras tratamos de preparar el ingrediente que va a continuación, leer el próximo pasaje del piano mientras escuchamos lo que toca el violoncelo y ponemos las manos en posición para la próxima cascada de acordes, o simplemente alargamos la mano para coger la cerveza mientras hacemos zapping. Suceden tantas cosas a la vez, superpuestas en el tiempo, que sería difícil diseccionar todas sus dependencias, pero es posible silenciarlo todo y aislar un acto «único» para poder estudiarlo. Siéntese usted y quédese muy quieto durante un rato, trate de no pensar en nada, y luego, sin otra razón que el mero hecho de querer hacerlo, flexione la muñeca una vez. Una simple flexión, por favor, cada vez que, como se dice, se lo pida el espíritu.*

*** El equivalente de la expresión inglesa *When the spirit moves you* sería en realidad «Cuando el cuerpo te lo pida»; la necesidad de sustituir el «cuerpo» por el «espíritu» para conservar el sentido de la argumentación del autor da pie sin duda a una expresión nueva en castellano. (N. *delt.)***

Llamemos a su acto voluntario e intencional: ***¡flexión!*** Si monitorizamos su cerebro con un sistema de electrodos de superficie (basta con que estén en el cuero cabelludo, no será necesario insertarlos en el cerebro), veremos que la actividad cerebral previa a ***¡flexión!*** tiene un curso temporal definido y repetible, y también una forma determinada. Dura casi un segundo —entre 500 y 1.000 milisegundos— y termina cuando se mueve su muñeca (cosa que podemos detectar si hacemos que su muñeca interceptе un rayo de luz dirigido a una sencilla célula fotoeléctrica). El movimiento de la muñeca viene precedido por menos de 50 milisegundos de actividad en los nervios motores que descienden desde el córtex motor de su cerebro hasta los músculos de su antebrazo, pero lo han precedido 800 milisegundos —casi un segundo— de una actividad claramente detectable en su cerebro conocida como «potencial de disposición» [***readiness potential***] o PD (Kornhuber y Deecke, 1965). (Véase la figura 8.1.)

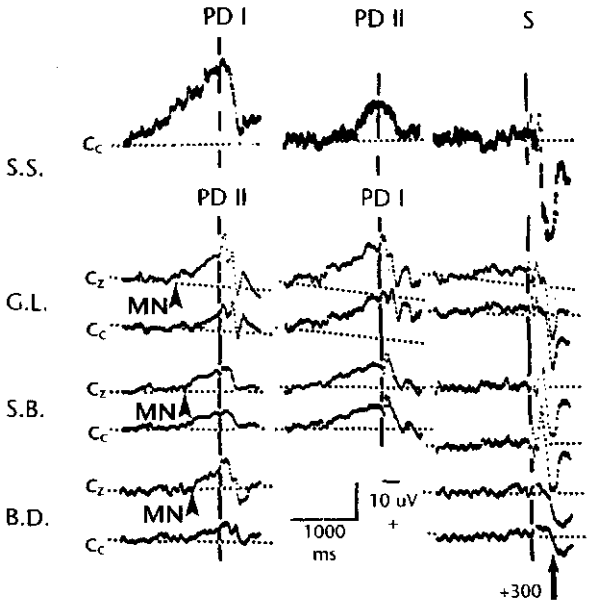

FIGURA 8.1. Registros de EEG de PD (Libet, 1999, pág. 46).

En algún punto de esos 1.000 milisegundos se encuentra el célebre «tiempo ***t»,*** el momento en el que ***usted*** decide conscientemente flexionar la muñeca. Benjamin Libet se propuso determinar exactamente cuándo era. Como este momento viene definido por sus propiedades subjetivas,

tuvo que dirigirse a *usted* para saber cuándo ocurre, para poder superponerlo luego a la serie de eventos objetivos que se producían en su cerebro. Diseñó un inteligente mecanismo para registrar ambas series, la subjetiva y la objetiva. Hacía que los sujetos miraran hacia un «reloj» con un punto que se movía rápidamente, como la segunda manecilla, pero a una velocidad considerablemente superior, una revolución cada 2,65 segundos, de modo que podía obtener lecturas de fracciones de segundos que luego podía contrastar con sus registros cronometrados de actividad cerebral (figura 8.2).

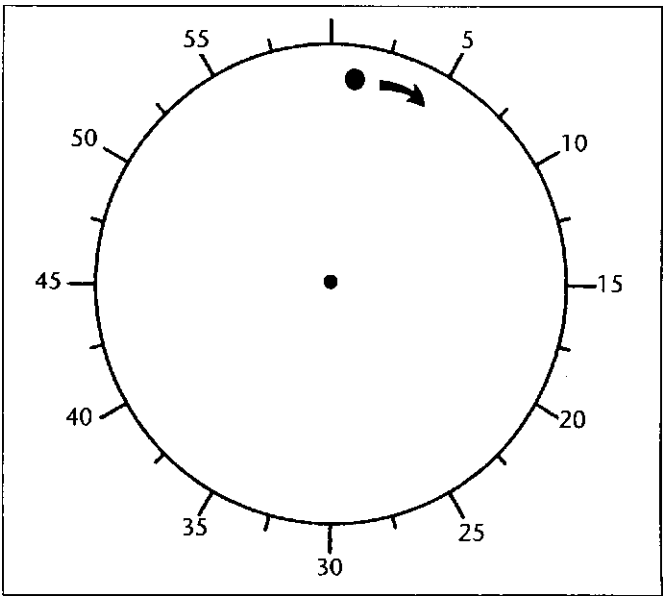

FIGURA 8.2. **Esfera de reloj usada por Libet (Libet,** 1999, **pág.** 48).

Libet pidió a sus sujetos que tomaran nota de la posición del punto en la esfera del reloj en el instante en que decidían flexionar la muñeca o eran conscientes del impulso o el deseo de flexionarla. Esta es la información que debían darle (más tarde, mucho después de haber hecho la flexión, sin necesidad de apresurarse a entregar su informe). Libet descubrió un agujero temporal o latencia entre el PD que medía en el cerebro de los sujetos y el *momento declarado de la decisión* de entre 300 y 500 milisegundos. Ese es el «vacío moral» del que habla Rachel Palmquist, y es enorme para los estándares neurocientíficos (en comparación, por ejem-

pío, con las inexactitudes e idiosincrasias observables en otros juicios de simultaneidad). No existe la menor duda de que, en esta circunstancia artificial, el PD es la causa de la flexión. El PD es un índice de predicción muy fiable de la flexión. Así pues, ¿dónde está el problema? Parece ser el siguiente: cuando usted *piensa* que está tomando una decisión, no hace más que contemplar pasivamente una especie de vídeo interno retrasado (con un inquietante retraso de 300 milisegundos) de la auténtica decisión que tuvo lugar inconscientemente en su cerebro un buen rato antes de que «se le ocurriera» flexionar la muñeca. Tal como dije en *La conciencia explicada-*.

> **No es que estemos «fuera de la cadena» (como dicen en la Casa Blanca), pero como accedemos a la información con retraso, lo único que podemos hacer es intervenir con «vetos» o «incentivos» de última hora. Como estoy más abajo del cuartel general (inconsciente), yo no tengo iniciativa real, nunca estoy presente en el nacimiento de un proyecto, pero ejerzo una moderada modulación ejecutiva sobre las políticas previamente dictadas que pasan por mi despacho (Dennett, 1991a, pág.** 164).

Pero en realidad yo exponía este punto de vista únicamente para demostrar su falsedad. A continuación decía: «Esta imagen es atractiva pero incoherente». Otras personas, sin embargo, no parecen ver la incoherencia. Tal como lo ha expresado el destacado (y bien vestido) neurocientífico Michael Gazzaniga: «Libet determinó que los potenciales cerebrales se activan 350 milisegundos antes de que tengamos la intención consciente de actuar. De modo que antes de que seamos conscientes de que estamos pensando en mover el brazo, nuestro cerebro ya está trabajando para realizar el movimiento» (Gazzaniga, 1998, pág. 73). William Calvin, otro notable (y sin duda debidamente trajeado) neurocientífico se expresa con mayor prudencia:

> **Mi colega neurofisiólogo Ben Libet ha demostrado, para consternación de todos, que la actividad cerebral asociada a la preparación del movimiento (algo llamado «potencial de disposición») [...] comienza un cuarto de segundo antes de que declaremos haber decidido movernos. Simplemente no éramos conscientes de nuestra decisión de movernos, pero ya estaba en curso (Calvin,** 1989, **págs.** 80-81).

Y el propio Libet ha resumido recientemente su propia interpretación del fenómeno en los siguientes términos:

La iniciación del acto voluntario y libre parece tener lugar en el cerebro de un modo inconsciente, antes de que la persona sepa conscientemente que quiere actuar. ¿Qué papel le queda, pues, a la voluntad consciente en el acto voluntario? (véase Libet, 1985). Para responder a esto es preciso reconocer que la voluntad consciente (V) se activa ciertamente unos 150 milisegundos antes de la activación del músculo, aunque sea posterior al inicio del PD. Un intervalo de 150 milisegundos sería tiempo suficiente para que la función consciente afectara al resultado final del proceso volitivo. (En realidad, sólo disponemos de 100 milisegundos a tal efecto. Los últimos 50 milisegundos previos a la activación del músculo es el tiempo que necesita el córtex motor primario para activar las células nerviosas motrices de la columna. Durante este tiempo el acto avanza hacia su culminación sin que el resto del córtex cerebral pueda hacer nada para detenerlo.) (Libet, 1999, pág. 49.)

Sólo una décima de segundo —100 milisegundos— durante la cual deben emitirse los vetos presidenciales. Tal como dijo una vez en broma el astuto (e impecablemente ataviado) neurocientífico Vilayanur Ramachandran: «Esto sugiere que nuestras mentes conscientes tal vez no sean libres de hacer cosas, sino más bien de no hacerlas» (Holmes, 1998, pág. 35). Odio mirarle el dentado a un caballo regalado, pero ciertamente quiero más libertad que eso. ¿Podemos encontrar algún fallo en el razonamiento que ha llevado a este distinguido grupo de neurocientíficos a esta conclusión?

La propuesta experimental de Libet es inusual y merece un examen cuidadoso. Estamos sentados tranquilamente, viendo cómo el punto del reloj da vueltas y vueltas, y esperamos a que, por ninguna razón excepto tal vez que nos estamos aburriendo, decidimos flexionar la muñeca: «Dejemos que el impulso de actuar aparezca por sí mismo en cualquier momento, sin ninguna preparación o concentración previa sobre cuándo hacerlo» (Libet y otros, 1983, pág. 625). Es importante que *no* apliquemos estrategias como la de decidir que flexionaremos la muñeca cada vez que la manecilla del reloj llegue a la posición de «las tres», puesto que entonces habríamos tomado una decisión («basada en nuestro libre albedrío») previa y no haríamos sino aplicarla de manera más o menos irreflexiva, en función de la información visual sobre la esfera del reloj. (Recordemos el caso de Martin Luther, que tomó su decisión hace largo tiempo, y ahora no puede hacer otra cosa.) ¿Cómo podemos estar seguros de que no estamos dejando que algún elemento de la esfera del reloj active nuestra «libre» decisión? Cada cual podrá pensar lo que quiera, pero por el momento supongamos que hemos conseguido seguir las instrucciones al menos hasta el siguiente punto: hasta donde somos capaces de decirlo, no esta-

mos «fijando» nuestra decisión a la posición del punto en el reloj, sino que sólo «registramos» en qué posición se encuentra la aguja cuando «se nos ocurre» flexionar la muñeca. Después de hacerlo, le decimos a Libet en qué posición estaba («El punto del reloj acababa de pasar las 10 cuando me decidí» o «El punto estaba abajo de todo, posición 30», o lo que sea), y estos datos previamente recogidos le permiten a Libet establecer con una precisión de milisegundos cuándo estaba el punto en aquella posición. Libet puede superponer entonces un registro temporal de nuestro flujo de conciencia (según nuestros informes posteriores) al de nuestra actividad cerebral, y eso determinará el momento en que tomamos conciencia de nuestra decisión, ¿correcto? Ésa es la premisa que subyace al experimento de Libet, pero no es tan inocente como podría parecer a primera vista.

Supongamos que en su caso Libet sabe que el potencial de disposición se activó en el milisegundo 6,810 del experimento, y que el punto del reloj estaba justo (según usted le dijo haber visto) en el milisegundo 7,005. ¿Cuántos milisegundos cabe esperar que debamos añadir a este número para llegar al momento en que usted fue consciente de ello? La luz llega desde la esfera del reloj hasta su ojo de forma casi instantánea, pero el recorrido de las señales desde la retina hasta el córtex estriado, pasando por los núcleos geniculados laterales, lleva entre 5 y 10 milisegundos (una fracción irrisoria de los 300 milisegundos de partida), pero ¿cuánto tiempo más tarda para llegar hasta *usted.*} (¿O es que se halla usted situado en el córtex estriado?) Las señales visuales deben ser procesadas antes de que lleguen a donde sea que deban llegar para que usted pueda tomar una decisión consciente de simultaneidad. El método de Libet presupone, en resumen, que podemos localizar la *intersección* de dos trayectorias:

- la emergencia a la conciencia de las señales referentes a la decisión de flexionar la muñeca,
- la emergencia a la conciencia de señales sucesivas referentes a la orientación del punto en la esfera del reloj,

de tal modo que dichos eventos se produzcan en paralelo, por así decirlo, en un lugar en el que pueda registrarse su simultaneidad. Como Libet quiere oírle hablar a *usted*, no a su córtex estriado, necesitamos saber dónde se encuentra *usted* ubicado en el cerebro antes de que podamos comenzar siquiera a interpretar los datos. Supongamos, por mor del argumento, que esto tiene sentido. Para ser justos y constructivos, dejemos a

un lado todas las versiones extravagantes de la suposición: Libet no supone que usted sea un homúnculo, con brazos y piernas, ojos y orejas, como el hombrecillo verde de la sala de control de la marioneta de tamaño humano de la morgue de *Men in Black: hombres de negro,* y no supone que usted sea una porción inmaterial de ectoplasma que fluye por su cerebro como una ameba fantasmática, ni que usted sea un ángel cuyas alas estarán dobladas hasta que sea llamado a subir volando al cielo. Debemos considerar una versión minimalista de la hipótesis, desembarazada de todos estos incómodos detalles: *usted* es «lo que sea que haga falta para poder experimentar la simultaneidad entre la decisión y la orientación del punto en la esfera del reloj». (Si necesitamos una imagen, podemos imaginarnos vagamente este «lo que sea» como algún nexo o núcleo de actividad cerebral, que puede cambiar de ubicación en diferentes condiciones, una onda cerebral con poderes cognitivos harto especiales. Véase la figura 8.3.) En tal caso quedan al menos tres posibilidades por explorar:

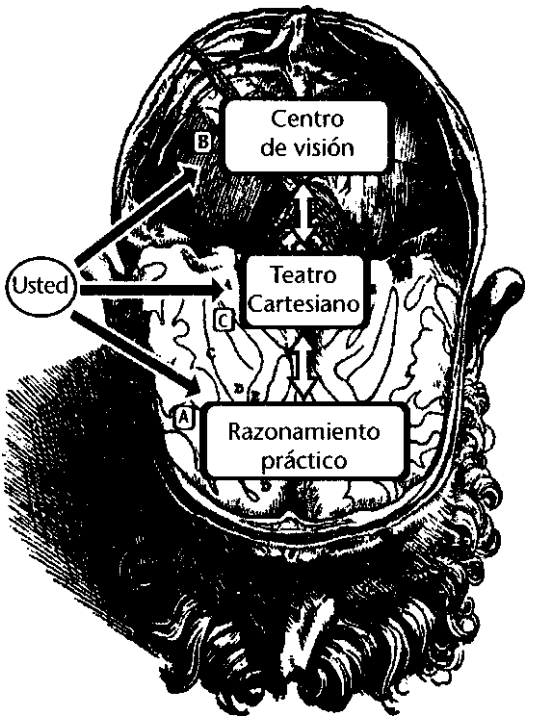

FIGURA 8.3. ¿Dónde se encuentra usted en su cerebro?

a) Usted está ocupado tomando una decisión libre en la *facultad de razonamiento práctico* (donde se toman todas las decisiones libres) y debe esperar a que la información visual le sea enviada desde el *centro de visión.* ¿Cuánto tiempo lleva eso? Si las presiones temporales no son críticas, puede suceder que el envío sea demasiado lento y la información ya esté desfasada para cuando llegue, como si recibiera el periódico de ayer.
b) Usted está ocupado mirando el reloj en el *centro de visión* y debe esperar a que la *facultad de razonamiento práctico* le envíe los resultados de su última decisión. ¿Cuánto lleva eso? De nuevo podríamos estar perdiendo un tiempo precioso con esta transmisión, ¿no es así?
c) Usted está donde siempre ha estado: en el *cuartel general* (conocido también como el Teatro Cartesiano), y tiene que esperar a que tanto el *centro de visión* como la *facultad de razonamiento práctico* envíen sus resultados respectivos a este señalado lugar, donde todo confluye y tiene lugar la conciencia. Si una de estas delegaciones está más lejos, o transmite más lentamente que la otra, estará usted expuesto a ilusiones de simultaneidad (si juzga usted la simultaneidad en función del momento de llegada al cuartel general, en lugar de confiar en algo así como un matasellos u otro tipo de marca temporal).

Plantear la cuestión de esta forma tan cruda ayuda —espero— a clarificar los problemas que presenta el modelo de Libet. ¿Cuál es la premisa no formulada que hay detrás de estas hipótesis? ¿Qué diferencia supondría para usted estar en uno u otro de estos lugares? La idea central, presumiblemente, es que usted sólo puede *actuar* desde el lugar en el que *está,* de modo que si no está *en* la facultad de razonamiento práctico cuando se toma la decisión, no es *usted* quien ha tomado esta decisión. En el mejor de los casos la habrá delegado. («Quiero estar *en* la facultad de razonamiento práctico. Después de todo, si no estoy *allí* cuando se toman las decisiones, las decisiones no serán mías. ¡Serán *suyas*!») Pero si usted se encuentra efectivamente allí, es posible que esté tan ocupado tomando la decisión que «pierda de vista» el trabajo que debe realizar el centro de visión, y nunca le llegue el resultado. De modo que tal vez sería mejor que fuera usted saltando entre el centro de visión y la facultad de razonamiento práctico. Pero si hiciera eso, cabe la posibilidad de que fuera consciente de la decisión *en el momento mismo en que la tomó,* pero tardara más de 300 milisegundos en ir hasta el centro de visión y obtener una imagen —tal vez llegara justo en el momento en que se recibía la imagen del pun-

to señalando hacia abajo—, con lo que habría juzgado mal la simultaneidad por haber perdido la cuenta del tiempo que tardó en ir de un sitio al otro. ¡Uf! Eso sería otra hipótesis más, llamémosla *Yo paseante*, que podría salvar la libertad, al demostrar que el agujero era una ilusión, después de todo. De acuerdo con esta hipótesis, usted decidió *conscientemente* flexionar la muñeca cuando esa parte de su cerebro decidió flexionarla (usted estaba allí, en ese mismo momento, dirigiendo el potencial de disposición en el momento de su creación), pero se equivocó al estimar el tiempo objetivo de la decisión según el reloj por culpa del tiempo que tardó en llegar al centro de visión y comprobar la última posición del punto en la esfera.

Si no le gusta esta hipótesis, ahí va otra que tal vez podría servir, basada en la alternativa (c), según la cual tanto el centro de visión como la facultad de razonamiento práctico serían desplazados fuera del cuartel general. Llamémoslo *Yo externo.* Usted ha delegado todas estas tareas a diversos subcontratistas, como se diría hoy en el mundo de los negocios, pero mantiene un control limitado sobre sus actividades desde su butaca del cuartel general, mediante el envío de órdenes y la recepción de los resultados, en un ciclo continuo de órdenes y respuestas. Si le piden una razón para no cenar fuera esta noche, usted manda a buscar una razón en la facultad de razonamiento práctico, y la recibe rápidamente: *estoy muy cansado y hay comida en la nevera que se echará a perder si no la comemos hoy.* ¿Cómo las encontró la facultad? Y ¿por qué en ese orden? ¿Qué operaciones realizó para generarlas? Usted no sabe nada de todo eso: simplemente sabe lo que usted ha pedido y reconoce que lo que ha recibido como respuesta se ajusta adecuadamente a su encargo. Si le preguntan qué hora es, usted envía la orden adecuada al centro de visión y éste responde con la última imagen de su reloj de muñeca, con algo de ayuda del *centro de control del movimiento de la muñeca*, pero no tiene ni idea de cómo se logró este éxito conjunto. Ante el problema de los retrasos temporales variables, establece un sistema de sellos temporales que funciona bien en la mayoría de los casos, pero falla en el contexto más bien artificial que propone Libet. Cuando se le pide que valore, desde su *desaventajada* posición en el cuartel general, en qué momento exacto emitió la orden de flexionar la muñeca su facultad del razonamiento práctico (un juicio que usted debe realizar en función de los sellos temporales que encuentra en los informes que le llegan desde la facultad de razonamiento práctico y el centro de visión), se equivoca de informes. El hecho de tener que confiar en información de segunda mano (informes de los dos sub-

contratistas) hace que sea fácil equivocarse respecto a qué evento sucedió primero, o cuáles de ellos han sido simultáneos.

Una cosa que tiene a su favor esta hipótesis es que tales juicios de simultaneidad son actos antinaturales a menos que estén dirigidos a algún propósito concreto, como el intento de iniciar el *staccato* en el momento en que el director dé la entrada, o de darle a una bola rápida para mandarla por encima del lanzador. En dichos contextos naturales se pueden realizar auténticas hazañas de precisión temporal, pero es notoria la facilidad con la que se producen errores o interferencias en los juicios aislados de simultaneidad «intermodales» (responder a preguntas tales como ¿qué fue primero, la señal luminosa o el pitido, o fueron simultáneos?). Lo que se presenta subjetivamente como simultáneo puede variar por entero en función de cómo realizamos un juicio, de qué uso le queramos dar. De modo que si usted realiza sus juicios de simultaneidad desde una posición tan poco aventajada, sin ningún contexto natural que aporte una razón para el juicio, es muy posible que ordene a la facultad de razonamiento práctico que emita una decisión y simplemente traspapele el informe final, de modo que se equivoque al estimar que ha sido simultáneo a la percepción del centro de visión de que el punto estaba en la posición del 30. Pero tal vez esta hipótesis no le resulte atractiva, puesto que de acuerdo con ella *usted* no está de hecho *presente* en la facultad de razonamiento práctico cuando toma la decisión.

Todavía queda otra hipótesis, que le devuelve a usted allí donde está (o estaba) la acción: la *tinta de secado lento.* Cuando usted toma una decisión, de manera consciente, *en* la facultad de razonamiento práctico (y usted está ahí, en el meollo del asunto), «escribe el resultado» con una tinta que es de secado lento: usted puede comenzar a actuar inmediatamente de acuerdo con ella, pero no podrá compararla con lo que ocurre en el campo de la visión hasta que la tinta se seque (al cabo de unos 300 milisegundos). (Esta hipótesis está inspirada en otro experimento de Libet que trato en *La conciencia explicada* [Dennett, 1991a], relativo a la «referencia hacia atrás en el tiempo» de la conciencia.) Según esta hipótesis, *usted* es quien decide ejecutar la *¡flexión!* exactamente cuando se inicia el PD en su cerebro, sin ningún retraso, pero no compara la decisión consciente con el resultado procedente del centro de visión hasta que han pasado unos buenos 300 milisegundos más, el tiempo que tarda su decisión en secarse antes de entrar en la sala de comparaciones.

Y si no le gusta tampoco esta hipótesis, todavía se podrían considerar otras, incluidas, por supuesto, todas las hipótesis que *no* «salvan la liber-

tad» porque tienden a confirmar la idea de Libet: que en el curso normal de la toma de una decisión moral, *usted* dispone como máximo de unos 100 milisegundos para emitir su veto o introducir cualquier otra corrección en las decisiones tomadas previamente (y en otro lugar) de manera inconsciente. Y, sin embargo, ¿no sería mejor descartar en bloque todas estas pobres hipótesis, sobre la base de que constituyen simplificaciones extremas y nada realistas de todo cuanto sabemos sobre el procesamiento de toma de decisiones en el cerebro? Claro que podríamos hacerlo, y es incluso lo que deberíamos hacer. Pero al hacerlo no descartamos únicamente las fantasiosas hipótesis que podrían «salvar la libertad» frente a los datos de Libet; también debemos descartar la propia hipótesis de Libet y todas las demás que pretenden demostrar que sólo tenemos «libertad de no hacer». Esta hipótesis depende tanto como las que acabamos de esbozar de que nos tomemos en serio la idea de que usted está limitado a los materiales a los que tiene acceso en una región particular del cerebro. ¿Cómo es eso? Consideremos la idea de que exista una ventana de oportunidad muy estrecha para emitir el veto. Libet presupone tácitamente que usted no puede considerar seriamente si quiere o no vetar algo hasta que es consciente de lo que usted podría querer vetar, y debe esperar 300 milisegundos o más para que esto ocurra, lo que le deja sólo 100 milisegundos para «actuar»: «Esto proporciona un período durante el cual la función consciente estaría en posición de determinar si el proceso volitivo llegará o no a completarse» (Libet, 1993, pág. 134). La «función consciente» espera, en el Teatro Cartesiano, hasta que llega la información, y *sólo entonces* tiene acceso a ella y puede ponerse a pensar en lo que puede hacer respecto a ella, si debe vetarla, etc. Pero, ¿por qué no podría haber estado pensando hace medio segundo («inconscientemente») en la posibilidad de flexionar la muñeca? Libet presupone necesariamente que el cerebro es capaz de resolver los detalles de la implementación del acto en el curso de dicho período de tiempo, pero sólo una «función consciente» tiene el talento requerido para examinar los pros y los contras de una decisión de veto.

En realidad, el propio Libet se da cuenta de este problema en cierto punto y lo aborda directamente: «No se excluye la posibilidad de que algunos de los factores en los que se *basa* la decisión de veto (control) se desarrollen a través de procesos inconscientes previos al veto» (Libet, 1999, pág. 51). Pero si no se excluye tal posibilidad, la conclusión que deberían sacar Libet y los demás es que el «agujero» de los 300 milisegundos *no* ha quedado en absoluto demostrado. Después de todo, sabemos que en circunstancias normales el cerebro inicia su labor de discriminación y eva-

luación tan pronto como recibe los estímulos, y trabaja simultáneamente en numerosos proyectos, lo que nos permite responder inteligentemente y a tiempo a muchos requerimientos distintos, sin tener que ponerlos en lista de espera para que pasen por el molinete de la conciencia antes de que comience su evaluación. Patricia Churchland (1981) demostró este punto con un sencillo experimento en el que pedía a los sujetos que respondieran conscientemente (¿cómo si no?) a una señal luminosa. Su tiempo total de respuesta era alrededor de 350 milisegundos. La reacción de Libet al descubrimiento de Churchland fue insistir en que tal respuesta se había iniciado inconscientemente: «La capacidad de detectar un estímulo y reaccionar resueltamente ante él, o estar psicológicamente influido por el mismo, sin que el sujeto reconozca ninguna conciencia del estímulo, es una posibilidad ampliamente aceptada» (Libet, 1981, pág. 188). Pero eso concede justamente el punto en discusión: usted puede reaccionar ante —o estar psicológicamente influido por— la decisión de flexionar la muñeca mucho antes de que dicha decisión «emerja a la conciencia». A pesar de todos los experimentos de Libet, usted podría tener acceso *óptimo* en todo momento a las decisiones que toma. Es decir, podría ser que todas las partes de su yo competentes para desempeñar alguna función en las decisiones que le corresponda a usted tomar reciban todo cuanto necesitan para hacer su trabajo en el menor tiempo posible. (¿Qué otra cosa podría preocuparle cuando se pregunta si recibe la información a tiempo para introducir los cambios que quiera?)

Los datos de Libet sí descartan una hipótesis, que tal vez hubiera sido nuestra favorita: el *Yo autocontenido,* según la cual *todas* las rutinas del cerebro se hallan concentradas en una única localización compacta, donde todo confluye en un mismo punto: la visión, el oído, las decisiones, los juicios de simultaneidad... Teniéndolo todo tan a mano, no se plantearía ningún problema temporal: una persona, un alma, podría instalarse allí tranquilamente y tomar decisiones libres y responsables, y ser simultáneamente consciente de éstas y de todo lo que ocurre en su conciencia en aquel momento. Pero no hay tal lugar en el cerebro. Tal como nunca me canso de señalar, todo el trabajo que realiza el imaginario homúnculo del Teatro Cartesiano debe ser dividido y repartido en el espacio y *en el tiempo* entre diversas instancias cerebrales. Vuelve a ser momento de repetir mi irónico lema: si uno se hace lo bastante pequeño, puede llegar a externalizarlo prácticamente todo.

El cerebro necesita tiempo para procesar los estímulos, y la cantidad de tiempo depende de qué información utilice y para qué propósitos. Un

jugador de tenis de élite puede devolver un servicio del rival en unos 100 milisegundos. Un servicio de Venus Williams (a una media de 200 kilómetros por hora) puede cruzar los 23 metros que separan las dos líneas de fondo en menos de 450 milisegundos, sólo 50 milisegundos más de los que tardó el servicio más rápido registrado en la historia (el de Greg Rusedski, a 236 kilómetros por hora de velocidad inicial). Y como la precisión temporal y el diseño del resto dependen decisivamente de la información visual (si usted se inclina a dudarlo, trate de restar con los ojos vendados), el cerebro debe ser capaz de recoger la información visual y canalizarla hacia un uso muy preciso en muy escaso tiempo. Tal como demostró Churchland, un sujeto normal tarda unos 350 milisegundos en apretar un botón si se le pide que indique cuándo ve una señal luminosa. Ahora bien, todas éstas son respuestas conscientes, voluntarias e intencionales *(¿o* no?), y no experimentan ningún retraso de 300 o 500 milisegundos. Por supuesto, tanto el jugador de tenis como el sujeto del experimento deben haber decidido antes (de manera libre y consciente) que iban a ajustar sus respuestas a ciertas condiciones particulares. Son casos como el de Luther, a escala reducida. El jugador de tenis se compromete previamente con un plan sencillo y luego permite que los «reflejos» ejecuten su acto intencional. (Puede conservar *cierta* condicionalidad, del tipo de *Si alto al revés* ENTONCES *globo defensivo* SI NO *cortada al fondo de la pista.* Lo que hace el jugador, en efecto, es convertirse temporalmente en una máquina de situación-acción.) Y cuando decidimos cooperar con el experimentador apretando el botón cada vez que aparezca la señal luminosa, estamos haciendo lo mismo: nos sentamos, ponemos el piloto automático y dejamos que se lleve a cabo la decisión. «No podía hacer otra cosa», podríamos decir. «Como no había tiempo de considerar ni de reflexionar, hice todas mis consideraciones por adelantado, cuando disponía de todo el tiempo que quisiera, para que cuando llegara el momento pudiera actuar sin necesidad de pensar.»

Esto es algo que hacemos continuamente. Nuestras vidas están llenas de decisiones que llevar a cabo cuando llegue el momento, compromisos revisables con estrategias y actitudes que configurarán unas respuestas que deberemos emitir con demasiada rapidez en el curso de la acción como para poder considerarlas reflexivamente. Somos los creadores y los ejecutores de estas políticas, aunque las compilemos en partes que sólo podemos controlar y supervisar indirectamente. El hecho de que podamos tocar en una banda musical, por ejemplo, demuestra que nuestros cerebros son capaces de realizar muchas tareas a la vez siguiendo unos pa-

trones temporales de gran complejidad, y todo ello es algo deliberado, controlado e intencional. Las respuestas que damos en una conversación, las palabras mismas que nos decimos en silencio *a nosotros mismos* cuando reflexionamos sobre lo que vamos a hacer, son actos cuya preparación se remonta siempre al pasado. Lo que Libet descubrió no es que la conciencia apenas logra seguir el paso de las decisiones inconscientes, sino que la toma de decisiones conscientes requiere tiempo. Si debemos tomar una serie de decisiones conscientes, lo mejor que podemos hacer es reservar medio segundo de reflexión, más o menos, para cada una, pero si debemos ejercer nuestro control de manera más rápida nos veremos obligados a compilar nuestras decisiones en una rutina que descarte buena parte del procesamiento requerido para una decisión consciente única. Libet recoge un sencillo experimento de Jensen (1979) que demuestra este efecto. Jensen pidió a los sujetos que presionaran un botón tan pronto como fueran conscientes de una señal luminosa, igual que había hecho Patricia Churchland, y obtuvo unos resultados consistentes con los suyos (en realidad, los tiempos de reacción de sus sujetos eran bastante más rápidos: una media de 250 milisegundos). Luego pidió a sus sujetos que retrasaran la activación del botón sólo un poco, *tan poco como fuera posible.* Tuvieron que añadir otros larguísimos 300 milisegundos a su tiempo de respuesta. El cerebro tiene trucos para evitar esos retrasos bajo ciertas circunstancias, como por ejemplo cuando buscamos ciertos objetos en un determinado escenario y disponemos de poco tiempo. Al buscar un objeto en concreto, por ejemplo, el cerebro opta a veces por dejarse ir; realiza un repaso visual al azar sobre el conjunto de la muestra, aunque pudiera hacer una búsqueda metódica «más eficiente». La atención salta más rápidamente de un objeto a otro cuando se la deja a su aire, puesto que «la atención es rápida, mientras que la voluntad es lenta» (Wolfe, Alvarez y Horowitz, 2000).

Habitualmente estos trucos encajan sin problemas entre sí y son incorporados a las tareas de control del cerebro sobre su propia actividad, pero en circunstancias artificiales (como las que astutamente diseñan los experimentadores) pueden ponerse de manifiesto. Por ejemplo, cuando el cerebro ejecuta una decisión de actuar (en el momento en que se aprecia el PD) proyecta una serie de anticipaciones —produce un pequeño futuro— de lo que debería ocurrir. Si lo que ocurre a continuación es alterado artificialmente —por ejemplo se acelera o se retrasa—, ello genera violaciones de dichas anticipaciones e indica que algo va mal. Pero es posible que el cerebro no pueda generar la interpretación correcta de lo que ha ocurrido en este

contexto desconocido. En *La conciencia explicada* (Dennett, 1991a, págs. 167-168), describí un experimento temprano que ilustraba esta posibilidad, y lo bauticé como el carrusel precognitivo de Grey Walter. A principios de la década de 1960, el eminente neurocirujano y pionero de la robótica Grey Walter quiso aprovechar la circunstancia de disponer de una serie de pacientes epilépticos con unos electrodos implantados en sus áreas motrices. Conectó los cables de los electrodos a un proyector de diapositivas con carrusel, de modo que cada vez que los pacientes decidían (improvisadamente, cuando el espíritu se lo pedía) pasar a la siguiente diapositiva, la actividad detectada en la región motriz disparaba directamente el avance del carrusel. El botón que pulsaban los pacientes era de pega, es decir, no iba conectado a nada. El efecto, según Walter, era impactante: a los pacientes les parecía que cada vez que «iban a» presionar el botón, pero siempre *antes* de que llegaran a decidirlo, el proyector les leía la mente y se les adelantaba.[2] La percepción prevista de un cambio en la diapositiva se veía «escamoteada» por una percepción ligeramente anterior del mismo cambio, lo que suscitaba en ellos una firme y siniestra impresión de que algo extraño estaba ocurriendo; el proyector les estaba leyendo la mente. En cierto sentido eso era exactamente lo que ocurría, pero no se trataba de que supiera sus decisiones antes de que ellos fueran conscientes de ellas: sólo estaba «leyendo» y ejecutando sus decisiones conscientes antes de que sus propios

2. Grey Walter describió este experimento en una charla a la que yo asistí en Oxford en 1963 o 1964. Por lo que yo sé, no llegó a publicarse ningún artículo al respecto. Yo mismo y otros académicos hemos tratado de localizarlo, sin éxito, y algunos —incluido Wegner— han expresado la sospecha de que Grey Walter nos estaba tomando el pelo aquel día en Oxford. Tal vez sea así, pero mi hipótesis personal es que Grey pudo decidir no publicarlo porque la validez ética del experimento era discutible, incluso para los estándares del momento: sus pacientes llevaban a menudo implantes que sobresalían del cráneo durante meses y meses, un régimen al que probablemente no hubieran dado su asentimiento si no hubieran pensado que formaba parte de un tratamiento que podía mejorar su epilepsia; según recuerdo, sin embargo, sus repetidas visitas al Instituto Burden de Grey Walter no tenían otro fin que el de participar en experimentos carentes de beneficios terapéuticos plausibles para ellos. (En todo caso, el experimento podría reproducirse en la actualidad con sujetos normales y en condiciones no invasivas, gracias a avanzados sistemas de análisis ultrarrápido de señales obtenidas de un escáner MEG o de electrodos colocados sobre el cuero cabelludo. El principal obstáculo técnico no es tanto obtener los datos como procesarlos en tiempo real con la rapidez suficiente como para generar el efecto de anticipación. Aunque no tengo noticia de ninguna reproducción publicada del experimento —o de fracasos en este sentido—, mi predicción es que cualquiera que se tome la molestia de realizar dicho test y las variantes del mismo que propongo en la página 168 de *La conciencia explicada* encontrará el efecto previsto.)

músculos del brazo pudieran «leerlas» y ejecutarlas. Imaginemos que ponemos una fotografía en un sobre y la enviamos a un amigo (por correo ordinario), y supongamos que la carta es inmediatamente interceptada por un ladrón de correo que, para hacer una gamberrada, escanea nuestra fotografía y se la envía por e-mail a nuestro amigo minutos después de que echemos el sobre al correo. Media hora después de haber enviado la fotografía, nuestro amigo nos llama y se maravilla ante los detalles de la imagen. Nosotros preveíamos exactamente esa llamada, pero no antes de dos o tres días. El hecho nos parecería inquietante, cuando menos, y tal vez nos sentiríamos tentados de llegar a la falsa conclusión de que debemos de haber enviado la carta mucho antes de que fuéramos conscientes de haberlo hecho: ¿no nos habremos levantado sonámbulos de la cama un día de éstos?

El desfase de 300 milisegundos de los sujetos de Libet podría obedecer a una confusión parecida. Cuando realizamos una acción intencional, normalmente la supervisamos visualmente (también mediante el oído y el tacto, por supuesto) para asegurarnos de que se desarrolla según lo previsto. La coordinación entre el ojo y la mano es posible gracias a una estrecha coordinación de sistemas sensoriales y motores. Supongamos que estoy mecanografiando intencionalmente las palabras «flexionar la muñeca» y deseo revisar mi producción en busca de errores tipográficos. Como las instrucciones motrices tardan cierto tiempo en ejecutarse, mi cerebro *no* debería comparar la instrucción motriz *actual* con la percepción visual *actual,* puesto que para cuando vea la palabra «flexionar» en la pantalla, mi cerebro ya estará enviando a mis músculos la instrucción de *escribir «muñeca».* Para realizar un control visual eficiente, mi cerebro debería retener la instrucción anterior (*escribe «giro»)* durante un cierto tiempo (¿tinta de secado lento?). Si dicho hábito estuviera lo bastante arraigado (¿y por qué no habría de estarlo?), debería interferir con el intento de realizar el acto más bien antinatural de determinar el momento de la decisión y no el de la acción ejecutada. La única forma de que los datos de Libet pudieran prestar algún tipo de apoyo al inquietante agujero de 300 milisegundos es suponer que el juicio de simultaneidad que propone no se vea distorsionado por ningún hábito de este tipo, pero tenemos buenas razones para creer que no es así, por lo que debemos considerar el agujero como un producto de una teoría mal construida y no como un auténtico descubrimiento.

Cuando dejamos atrás el cuello de botella cartesiano, y con él nuestro compromiso con el ideal de un mítico tiempo *t* (el instante en que se produce la decisión consciente), el descubrimiento de Libet de la ventana de 100 milisegundos para emitir el veto se esfuma por completo. Vemos en-

tonces que nuestra libertad de decisión, como todas nuestras demás capacidades mentales, debe extenderse en el tiempo, no medirse por instantes. En cuanto distribuimos por el cerebro, tanto en el espacio como en el tiempo, el trabajo que antes realizaba el homúnculo (en este caso, la toma de la decisión, la comprobación del reloj y el juicio de simultaneidad), también tenemos que distribuir la agencia moral. No estamos fuera de la cadena; *somos* la cadena. Somos tan grandes como eso. No somos un punto sin extensión. Lo que hacemos y lo que somos *incorpora* todas esas cosas que ocurren: no es nada distinto de ellas. En cuanto comience a verse a usted mismo desde esta perspectiva, podrá descartar el concepto antes tan persuasivo de una actividad mental que tiene un *comienzo* /«consciente y que sólo posteriormente «emerge a la conciencia» (donde está *usted* esperando ansioso a tener acceso a ella). Esto no es más que una ilusión, puesto que muchas de nuestras reacciones ante esta actividad mental tienen su inicio en un momento anterior: hasta allí llegan nuestras «manos», tanto en el espacio como en el tiempo.[3]

LA PERSPECTIVA DE UN TELÉPATA

> *Ilusoria o no, la voluntad consciente es la guía de la persona hacia su responsabilidad moral en sus acciones.*
>
> DANIEL WEGNER, *Thelllusion ofConscious Will*

Si el modelo cartesiano de Libet sobre la toma consciente de decisiones es demasiado simple, ¿qué aspecto tendría un modelo mejor? El modelo de Daniel Wegner tiene la curiosa virtud de estar a medio camino en la dirección correcta. Es todavía demasiado cartesiano, demasiado dependiente de la seductora metáfora del «lugar donde estoy en el cerebro», lo cual no hace más que ilustrar el poderoso atractivo de esta idea. Resulta

3. Un comentarista de Libet, Sean Gallagher, se acerca mucho a esta conclusión: «Pienso que este problema puede resolverse si dejamos de concebir la decisión libre como un acto momentáneo. En cuanto comprendemos que la deliberación y la decisión son procesos que se extienden en el tiempo, aunque sea, en algunos casos, en intervalos muy cortos de tiempo, se abre un gran margen para componentes conscientes que sean más que accesorios incorporados *a posteriori*» (Gallagher, 1998). (Pero luego pasa a decir que si la retroalimentación es inconsciente, será «determinista», pero que si es consciente, no lo será. Es difícil superar el pensamiento cartesiano.)

muy difícil, realmente, describir la fenomenología inmediata de la toma de decisiones en otros términos, de modo que si tratamos de orientarnos desde el refugio que ha construido Wegner a medio camino, tal vez podamos ver mejor cómo completar nuestra salida del Teatro Cartesiano.

Todo el mundo sabe lo que se supone que es capaz de hacer un telépata; pues bien, Wegner es un consumado telépata, pero no tanto para leer como para *escribir* en otras mentes. Wegner ha descubierto la manera de diseñar acciones intencionales e imponerlas a otras personas, de modo que piensen que han decidido realizarlas por sí mismas. Existe una pequeña industria artesanal en el mundillo filosófico de los estudios sobre la libertad dedicada al estudio de toda clase de experimentos mentales relativos a diversos personajes imaginarios capaces de escribir en la mente de otros, como por ejemplo unos nefandos neurocirujanos empeñados en implantar sistemas de control remoto en los cerebros de sus víctimas; sin embargo, los experimentos reales sobre la escritura de mentes introducen algunas novedades que poseen, en mi opinión, mayor interés filosófico.

¿Cómo puede alguien escribir intenciones en la mente de otro? ¿No tenemos cada uno de nosotros un «acceso privilegiado» a nuestras propias elecciones y decisiones? En realidad, no. Uno de los temas centrales en la obra de Wegner es la demostración, por diversas vías, de que nuestro conocimiento de la relación entre nuestros pensamientos y nuestras acciones (y entre unos pensamientos y otros pensamientos) sólo tiene el «privilegio» de la familiaridad. Si yo sé mejor que tú lo que me dispongo a hacer, es sólo porque paso más tiempo conmigo mismo que tú. Pero si alguien introduce subrepticiamente en mi flujo de conciencia las bases para una falsa creencia, puede hacerme pensar que estoy tomando decisiones «libres» cuando en realidad es él quien controla mis acciones. La técnica básica es bien conocida por los magos desde hace siglos: los magos la llaman actualmente *sugestión psicológica*, y es notablemente efectiva en manos competentes. Le damos a la víctima razones diversas para pensar que él y sólo él es responsable de una decisión que nosotros queremos que tome, y cae en la trampa. O bien podemos engañarle en sentido contrario, y hacerle pensar que no es responsable de un efecto que en realidad está produciendo él mismo (por ejemplo, un mensaje enviado por los «espíritus» sobre un tablero ouija).

Wegner ha adaptado el principio del tablero ouija y las técnicas de los magos a un laboratorio, y ha obtenido algunos resultados notables. Los sujetos de sus experimentos son inducidos sistemáticamente a atribuirse erróneamente decisiones que en realidad han sido tomadas por otra per-

sona. La razón de que se les pueda engañar, tal como David Hume observó con elocuencia hace varios siglos, es que las relaciones causales son imposibles de *percibir*. No podemos verlas cuando se dan fuera de nosotros, y no podemos descubrirlas por introspección cuando ocurren en nuestro interior. Lo que la gente percibe es que primero ocurre una cosa y luego otra, y caen en el truco de magia de Wegner por las mismas razones por las que caemos en la magia ordinaria: estamos demasiado predispuestos a interpretar, a «ver» cómo unas cosas causan otras, cuando en realidad tanto la «causa» como el «efecto» son efectos de una compleja maquinaria que se mantiene oculta para nosotros (literalmente entre bastidores). Wegner demuestra que no tenemos nada parecido a un acceso directo a las causas y los efectos de nuestras decisiones e intenciones, sino que debemos realizar inferencias (y hacerlo deprisa y sin mucha fanfarria lógica). En realidad somos muy buenos en esto; las inferencias que realizamos son casi siempre inferencias hacia la mejor explicación de la secuencia que hemos experimentado, excepto cuando un astuto manipulador ha puesto algunas premisas engañosas en el escenario.

Nótese cómo la introducción de la cuestión del acceso privilegiado nos sitúa automáticamente en una pendiente peligrosa que nos arrastra hacia el Teatro Cartesiano: hay cosas que ocurren dentro de mí de las que yo no sé nada, y hay otras cosas que conozco «directamente» (se *me* presentan de algún modo, esté donde esté). En lugar de luchar contra esta pendiente interpretativa, Wegner se permite caer en el escenario cartesiano completo cada vez que sirve a sus propósitos: «No podemos saber (y menos aún controlar) el inmenso número de influencias mecánicas que dirigen nuestro comportamiento, porque habitamos una máquina extraordinariamente compleja» (Wegner, 2002, pág. 27). Estas máquinas que habitamos simplifican las cosas para nosotros: «La experiencia de la voluntad es, pues, la forma que tienen nuestras mentes de representar sus operaciones ante nosotros, no su funcionamiento real» (pág. 96). En otras palabras, obtenemos una imagen útil pero distorsionada de lo que ocurre en nuestro cerebro:

> La conveniencia peculiar del ser humano de tener pensamientos conscientes que prefiguren nuestras acciones nos otorga el privilegio de sentir que somos la causa de lo que hacemos. En realidad, son mecanismos inconscientes e inescrutables los que generan tanto el pensamiento consciente sobre la acción como la acción misma, y también producen la sensación de voluntad que experimentamos al percibir el pensamiento como causa de la acción. Así pues, si bien nuestros pensamientos pueden tener conexiones causales profundas, importantes e inconscientes con nuestras acciones, la experiencia

de la voluntad consciente emerge de un proceso que interpreta dichas conexiones, no de las conexiones mismas (Wegner, 2002, pág. 98).

¿Quién o qué es este «nosotros» que habita en nuestro cerebro? Es un comentarista o un intérprete con un acceso limitado a la maquinaria efectiva, más parecido a un secretario de prensa que a un presidente o un jefe. Y esta imagen nos lleva directamente a la visión de Libet de la «voluntad consciente» como algo que está fuera de la cadena.

La conciencia y la acción parecen jugar al juego del gato y el ratón a lo largo del tiempo. Por más que podamos tener acceso consciente a panoramas muy completos de nuestras acciones antes de que éstas tengan lugar, es como si la mente consciente quedara desconectada de ellas. Un microanálisis del intervalo temporal previo y posterior a la acción indica que la conciencia entra y sale de escena y *en realidad no hace nada* [la cursiva es mía]. Las investigaciones de Libet, por ejemplo, sugieren que cuando llegamos al instante en que se produce una acción espontánea, la experiencia de querer conscientemente la acción se produce únicamente después de las señales de PD de que los procesos cerebrales han comenzado a generar la acción (y probablemente también la intención y la experiencia de voluntad consciente) (Wegner, 2002, pág. 59).

UN YO PROPIO

Nunca dejan de sorprenderme todos estos tratos que tengo conmigo mismo. Primero he acordado un principio conmigo mismo, ahora estoy defendiendo una tesis delante de mí mismo, y debato mis propios sentimientos e intenciones conmigo. ¿Quién es este yo, este compañero interno fantasma, con el que entro en todos estos comercios? (Me pregunto a mí mismo.)

MICHAEL FRAYN, *Headlong*

Los filósofos y los psicólogos acostumbran a hablar de un órgano unificador llamado el «yo» que puede «ser», según los casos, autónomo, dividido, individualizado, frágil, bien delimitado, etc., aunque no es un órgano que deba existir como tal.

GEORGE AINSLIE, *Breakdown ofWill*

Una acción voluntaria es algo que una persona puede hacer cuando se lo piden.

DANIEL WEGNER, *The Illusion ofConscious Will*

Así pues, según Wegner, «la conciencia [...] no hace nada en realidad», y es por ello que la voluntad consciente es, según proclama el título de su libro, una ilusión. Hay una vía de escape de esta visión, gracias a un ligero cambio de perspectiva que se encuentra de hecho implícito en la obra de Wegner. La conciencia tiene mucho trabajo que hacer, pero sus actividades parecen desaparecer cuando nos preguntamos a nosotros mismos qué está haciendo *ahora mismo* (en el tiempo *i)*. Y puesto que en cada momento concreto «no hace nada en realidad», puede parecer que es un acompañamiento meramente epifenoménico, un extra. Una perspectiva evolucionista muestra por qué esto no es así.

Uno de los fenómenos que Wegner señala en apoyo de su nueva concepción es la «automaticidad ideomotriz». Es el nombre de un fenómeno familiar —aunque siempre inquietante— que consiste en pensar en una cosa y que se produzca una acción corporal relacionada con la misma, sin que dicha acción sea una acción intencional. Por ejemplo, puede darse el caso de que descubramos, avergonzados, que hemos traicionado un pensamiento erótico secreto con un revelador gesto de la mano que no teníamos intención de hacer. En un caso como éste no somos conscientes de la relación causal entre el pensamiento y el acto, pero ésta existe, tan cierta como la relación causal que hay entre el aroma de la comida y la salivación. El rasgo principal de las acciones ideomotrices es que nos pasan desapercibidas a las personas (podría decirse que tenemos un acceso *desaventajado* a ellas). Es como si en nuestras mentes habitualmente transparentes hubiera ciertas barreras o cortinas, tras las cuales las cadenas causales siguen su curso sin que nosotros seamos conscientes de ellas, y producen sus efectos sin nuestro consentimiento. «Este ejército fantasma de acciones inconscientes plantea un serio desafío a la idea de un agente humano ideal. Los principales desmentidos de nuestro ideal de agencia consciente se producen cuando nos encontramos haciendo algo sin ningún pensamiento consciente de lo que estamos haciendo» (Wegner, 2002, pág. 157).

Para Descartes, la mente era perfectamente transparente a sí misma, nada ocurría fuera del escenario, e hizo falta más de un siglo de teoría y experimentación psicológica para erosionar su ideal de una introspección perfecta, que, según sabemos ahora, presenta la situación casi al revés de como es realmente. La conciencia de los flujos de actividad es la excepción, no la regla, y fue preciso que se dieran unas circunstancias más bien notables para que llegara a evolucionar. Las acciones ideomotrices son los restos fósiles de una época anterior en la que nuestros ancestros no esta-

ban tan informados como nosotros de lo que estaban haciendo. Tal como dice Wegner: «Más que necesitar una teoría especial para explicar la acción ideomotriz, tal vez sólo haga falta explicar por qué las acciones ideomotrices y los automatismos han eludido el mecanismo que produce la experiencia de la voluntad» (pág. 150).

En la mayoría de las especies que han existido, no hay ninguna necesidad de una causación «mental» y, por lo tanto, no ha evolucionado ninguna capacidad elaborada para la observación de uno mismo. En general, las causas funcionan perfectamente en la oscuridad, sin necesidad de que nadie las observe, y eso vale tanto para las causas que operan en los cerebros de los animales como en cualquier otro lugar. De modo que por más «cognitivas» que puedan ser las facultades de discriminación de un animal, la capacidad de sus respuestas para producir la selección del comportamiento adecuado no requiere que *nada o nadie* la experimente. El sistema nervioso de una simple criatura puede albergar un engranaje de vínculos de situación-acción de complejidad indefinida capaz de servir a sus muchas necesidades sin necesidad de ninguna supervisión ulterior. Es posible que sus acciones precisen la guía de un cierto control interno (específico para cada acción), para asegurarse, por ejemplo, de que cada tentativa depredadora llega hasta la presa, o para ponerse las bayas en la boca, o para guiar el delicado encaje con los órganos sexuales de un congénere del sexo opuesto, pero esos bucles de retroalimentación pueden ser tan aislados, tan locales, como los controles que ponen en marcha el sistema inmunológico cuando hay riesgo de infección o ajustan el ritmo de la respiración y los latidos del corazón al hacer ejercicio. (Esta es la verdad que hay detrás de la engañosa intuición de que los invertebrados podrían ser «robots» o «zombis», completamente carentes de mente, pero no así los animales «superiores, de sangre caliente».)

A medida que aumentan las opciones de comportamiento de las criaturas, sin embargo, sus mundos se van llenando de cosas, y la virtud de ser ordenado comienza a poder ser «apreciada» por la selección natural. Muchas criaturas han desarrollado sencillos comportamientos instintivos para lo que podría llamarse mejoras del hogar, preparación de caminos, escondites y puestos de vigilancia, además de otros aspectos de su vecindario, cuyo efecto es en general el de hacer más comprensible y manejable su entorno local. De modo parecido, cuando surge la necesidad, las criaturas desarrollan instintos para poner un poco de orden en su entorno más íntimo: su propio cerebro, donde crean también caminos y señales para su uso ulterior. El objetivo que persiguen inconscientemente

todos esos preparativos es que la criatura se oriente fácilmente en sí misma, y la cuestión de qué parte de estos trabajos de mejora doméstica son obra de la manipulación del propio individuo y qué parte es incorporada genéticamente queda abierta a un examen empírico. Alguno de estos caminos, o muchos de ellos, dio pie a las innovaciones que llevaron al surgimiento de criaturas capaces de considerar diferentes cursos de acción antes de comprometerse con alguno de ellos, y considerarlos sobre la base de algún tipo de proyección sobre el resultado que tendría cada uno. En el capítulo *5* consideramos el surgimiento de las máquinas de elección, capaces de evaluar los resultados probables de diferentes opciones antes de tomar una decisión. En la carrera de los cerebros por producir futuros útiles, ésta fue una innovación de primer orden respecto al *ciego* mecanismo de ensayo y error, puesto que, tal como dijo una vez Karl Popper, permite que algunas de nuestras hipótesis no lleguen a salir de casa. Dichas criaturas popperianas, tal como las llamo yo, ponen a prueba algunas de sus corazonadas en simulaciones informadas antes de arriesgarse a ponerlas en práctica en el mundo real, pero no tienen por qué comprender la razón que hay detrás de esta práctica innovadora para poder recoger sus beneficios. La apreciación de los efectos probables de acciones particulares se halla implícita en estas evaluaciones, pero la apreciación de los efectos de la contemplación misma constituye un nivel aún más elevado, y más optativo, de autocontrol. No tenemos por qué saber que somos una criatura popperiana para serlo. Al fin y al cabo, cualquier ordenador que juegue al ajedrez considera y descarta miles o incluso millones de jugadas posibles sobre la base de su resultado probable, y manifiestamente no es ningún agente consciente o autoconsciente. (Al menos no todavía: el futuro puede deparar robots conscientes y autoconscientes, los cuales entran sin duda alguna dentro de lo posible.)

¿Qué cambio se produjo en el mundo que animó la evolución de una implementación *menos inconsciente* del control popperiano del comportamiento? ¿Qué nueva complejidad ambiental favoreció las innovaciones en la estructura de control que lo hicieron posible? En una palabra, la comunicación. Sólo cuando una criatura comienza a desarrollar la actividad comunicativa, y en particular la comunicación de sus planes y acciones, puede esperarse que tenga alguna capacidad de contemplar no sólo los resultados de sus acciones, sino también sus evaluaciones previas y la formación de sus intenciones (McFarland, 1989). En ese punto, necesita un nivel de autocontrol que la mantenga informada de qué proyectos de si-

tuación-acción están en cola para ser ejecutados, o compiten para pasar a esta fase, y qué opciones está considerando la facultad de razonamiento práctico (si no nos parece una etiqueta demasiado grandilocuente para referirnos al terreno donde tiene lugar ahora la competición). ¿Cómo pudo surgir este nuevo talento? Estamos en condiciones de ofrecer una «Historia de así fue» que resalta los pasos cruciales del proceso.

Compárese la situación a la que se enfrentaban nuestros antepasados (y la Madre Naturaleza) con la situación a la que se enfrentaban los ingenieros de software que querían hacer los ordenadores más accesibles para el usuario. Los ordenadores son máquinas endiabladamente complejas, la mayoría de cuyos detalles son terriblemente enrevesados y, para la mayoría de propósitos, prescindibles. Los usuarios informáticos no necesitan estar informados de lo que pasa en todos sus circuitos, ni de la localización concreta de sus datos en el disco, etc., de modo que los diseñadores de software crearon una serie de simplificaciones —en algunos casos distorsiones inocentes— de la complicada verdad del asunto, astutamente diseñadas para adaptarse a —y potenciar— las capacidades previas de acción y percepción de los usuarios. Los actos de marcar y arrastrar con el ratón, los efectos sonoros y los iconos sobre el escritorio son los más evidentes y famosos, pero cualquiera que se tome la molestia de mirar más a fondo encontrará una infinidad de metáforas de este tipo que nos ayudan a interpretar lo que ocurre en el interior de la máquina, pero siempre al precio de simplificarlo. A medida que la gente comenzó a interactuar más con los ordenadores, se inventó muchos otros trucos, proyectos, objetivos y técnicas para usar y abusar de las competencias diseñadas para ellos por los ingenieros, los cuales volvieron luego a la plantilla de diseño para desarrollar nuevos refinamientos y mejoras, de los que luego usaron y abusaron otra vez los usuarios, un proceso coevolutivo que está lejos de terminar. La interfaz de usuario con la que interactuamos hoy era inimaginable cuando aparecieron los ordenadores, y en más de un sentido es la punta de un iceberg: no sólo se mantienen ocultos los detalles de lo que ocurre en el interior del ordenador, sino también los detalles de la historia de I+D, los falsos comienzos, las malas ideas que se esbozaron sin llegar jamás al gran público (así como las buenas ideas que sí lo hicieron y no llegaron a cuajar). Un proceso parecido de I+D fue el encargado de crear la interfaz de usuario entre los diferentes agentes parlantes, y se guió por similares razones (virtuales) y principios de diseño. Fue también un proceso de coevolución, en el que los comportamientos, las actitudes y los propósitos de las personas evolucionaron en respuesta a las nuevas capa-

cidades que iban descubriendo. Aprendieron que podían *hacer cosas con las palabras* que nunca habían podido hacer antes, y la belleza de todo el proceso consistía en que *tendía* a darles acceso desde el exterior a aquellos aspectos de sus complicados vecinos que más interesados estaban en ajustar (sin necesidad de saber nada de su sistema interno de control: el cerebro). Esos antepasados nuestros descubrieron clases enteras de comportamientos generativos para ajustar el comportamiento de los demás al suyo, y para supervisar y modular los intentos recíprocos de ajuste de sus propios controles de comportamiento por parte de otros (y, en caso necesario, resistirse a ellos).

La metáfora central de esta ilusión de usuario coevolutiva es el Yo, aparentemente ubicado en un cierto lugar del cerebro, el Teatro Cartesiano, y que proporciona una imagen limitada y metafórica de lo que ocurre en nuestros cerebros. Ofrece esta imagen a los demás, y *a nosotros mismos*. En realidad, no habríamos llegado nunca a existir —como yoes que «habitan una complicada maquinaria», según la elocuente expresión de Wegner— si no fuera por la evolución previa de interacciones sociales que requerían de cada animal humano que creara en su interior un subsistema diseñado para interactuar con los otros. Una vez creado, dicho subsistema podía interactuar también consigo mismo en diferentes momentos temporales. Hasta la llegada de los seres humanos, ningún agente del planeta disfrutó de la curiosa *no inconciencia* que nos caracteriza a nosotros en relación con los vínculos causales que se nos presentaron como más relevantes cuando comenzamos a hablar entre nosotros sobre lo que pretendíamos hacer.[4] Tal como lo expresa Wegner: «Las personas se convierten en lo que piensan que son, o en lo que descubren que los otros piensan que son, en un proceso de negociación que nunca se detiene» (Wegner, 2002, pág. 314).

Cuando los psicólogos y los neurocientíficos diseñan un nuevo entorno o paradigma experimental para estudiar sujetos no humanos como ratas, gatos, monos o delfines, a menudo tienen que dedicar decenas o incluso cientos de horas a entrenar a cada sujeto en sus nuevas tareas. Por ejemplo, se puede entrenar a un mono para que mire a la izquierda cada vez que vea un diseño de líneas paralelas moviéndose hacia arriba, y hacia la derecha cada vez que vea el mismo diseño moviéndose hacia abajo. Se puede entre-

4. Tal vez algún filósofo quiera comparar mi «Historia de así fue» con el mito de Wilfrid Sellars (1963) sobre «nuestros antepasados ryleanos» y «Jones, el inventor de pensamientos». Mi deuda con Sellars debería ser patente para ellos.

nar a un delfín para que vaya a buscar un objeto que tenga el mismo aspecto (o sonido, para su sistema ecolocalizador) que un objeto que le muestra su entrenador. Todo este entrenamiento requiere tiempo y paciencia, tanto por parte el entrenador como del sujeto del experimento. A los sujetos humanos que participan en experimentos, en cambio, basta en general con decirles lo que se desea de ellos. Tras una breve sesión de preguntas y respuestas y unos minutos de práctica, los sujetos humanos seremos tan competentes en el nuevo contexto como pueda serlo cualquier otro agente. Por supuesto, debemos *comprender* las representaciones que se nos ofrecen en dichas sesiones, y lo que se nos pide debe estar integrado por actividades que entren dentro del espectro de cosas que podemos hacer. Eso es lo que Wegner quiere decir cuando identifica las acciones voluntarias como las cosas que podemos hacer cuando se nos pide que las hagamos. Si se nos pide que rebajemos nuestra presión sanguínea, ajustemos el ritmo de nuestros latidos o movamos las orejas, no estaremos en la misma disposición de cumplir con la demanda, aunque con la ayuda de cierto entrenamiento, no muy distinto del que reciben los animales de laboratorio, podremos incluir finalmente dichas proezas a nuestro repertorio de actos voluntarios.

Tal como me hizo notar Ray Jackendoff, el surgimiento del lenguaje trajo a la existencia un tipo de mente capaz de transformarse de manera casi instantánea en una máquina virtual en cierto modo distinta: asumir nuevos proyectos, obedecer a nuevas reglas y adoptar nuevas estrategias. Somos transformistas. Eso es lo que es una mente, y la distingue de un mero cerebro: el sistema de control de un transformista camaleónico, una máquina virtual para crear máquinas aún más virtuales. Los animales no humanos pueden realizar acciones más o menos voluntarias. El pájaro que vuela hacia donde quiere se dirige voluntariamente hacia aquí o hacia allá, moviendo sus alas por su voluntad, y lo hace sin necesidad de lenguaje. La distinción implícita en su anatomía entre lo que puede hacer voluntariamente (por acción de sus músculos estriados) y lo que ocurre autónomamente (por acción de los músculos lisos y controlado por el sistema nervioso autónomo) no es negociable. Nosotros hemos añadido otra capa a la capacidad del pájaro (y del simio, y del delfín) de decidir lo que va a hacer. No es una capa anatómica en el cerebro, sino una capa funcional, una capa virtual presente de algún modo en los microdetalles de la anatomía del cerebro: nos podemos pedir unos a otros que hagamos cosas, y nos podemos pedir a nosotros mismos que las hagamos. Y al menos en ciertas ocasiones satisfacemos prestamente estas demandas. Es cierto, también le podemos «pedir» a nuestro perro que haga diversos actos voluntarios, pero el perro no puede

preguntarnos a su vez por qué se lo pedimos. Un babuino macho puede «pedirle» ayuda para su aseo a una hembra que pase por su lado, pero ninguno de los dos puede debatir el resultado probable de la satisfacción de su demanda, lo que podría tener graves consecuencias para ambos, especialmente si el macho no es el macho alfa del grupo. Nosotros, los seres humanos, no sólo podemos hacer cosas cuando se nos pide que las hagamos; podemos responder a preguntas acerca de lo que estamos haciendo y acerca del porqué. Podemos participar en la práctica de pedir y dar razones.

Es esta capacidad de preguntar, que también podemos dirigir hacia nosotros mismos, la que da origen a la categoría especial de acciones que nos hacen distintos. Otros sistemas intencionales más simples actúan de un modo que resulta rígidamente predecible a partir de las creencias y los deseos que les atribuimos tomando como base nuestros estudios sobre su historia y sus necesidades, sus capacidades perceptivas y de comportamiento, pero algunas de nuestras acciones son, tal como insistía Robert Kane, autoformativas en un sentido moralmente relevante: son el resultado de las decisiones que hemos ido tomando en nuestro empeño por darnos sentido a nosotros mismos y a nuestras vidas (Coleman, 2001). En cuanto comenzamos a entrar en conversaciones sobre lo que hacemos, debemos tener una idea clara de eso que hacemos para disponer de respuestas para dichas preguntas. El lenguaje exige de nosotros que estemos atentos, pero también nos ayuda a estarlo, al facilitarnos la tarea de categorizar y (sobre)simplificar nuestras *agendas*. No podemos evitar convertirnos en psicólogos aficionados. Nicholas Humphrey y otros investigadores dicen de los monos y otras especies altamente sociales que son como unos *psicólogos naturales,* a causa del manifiesto talento e interés que dedican a interpretar el comportamiento de sus congéneres, pero como los simios nunca llegan a comparar sus respectivas experiencias, ni a debatir sobre atribuciones de motivos y creencias, a diferencia de lo que ocurre con los psicólogos académicos —y otros seres humanos—, su competencia como psicólogos nunca les obliga a emplear representaciones explícitas. En nuestro caso es diferente. Nosotros necesitamos tener algo que decir cuando nos preguntan qué se supone que estamos haciendo. Y cuando respondemos, nuestra autoridad para hacerlo es dudosa. El biólogo evolutivo William Hamilton expresó este punto con particular claridad al reflexionar sobre su propia incomodidad al reconocerlo:

> ¿Qué es lo que realmente quería yo en la vida? Mi propio yo consciente y aparentemente indivisible se estaba convirtiendo en algo bien distinto de lo

> que yo había imaginado, y no tenía por qué avergonzarme tanto de la lástima que sentía por mí mismo [...]. Era un embajador enviado al extranjero por cierta frágil coalición, portador de órdenes contradictorias emitidas por los volubles gobernantes de un imperio dividido [...]. Mientras escribo estas palabras, para que sea capaz siquiera de escribirlas, se pretende que soy una unidad que, en el fondo de mí mismo, sé que no existe (Hamilton, 1996, pág. 134).

Wegner tiene razón, pues, al identificar el yo que emerge en sus experimentos y en los de Libet como una especie de encargado de relaciones públicas, más como un portavoz que como un jefe, pero dichos experimentos son casos extremos diseñados para aislar unos factores que se presentan habitualmente integrados, y no tenemos por qué identificarnos tan estrechamente con este yo temporalmente aislado. (Si uno se hace lo bastante pequeño...) Wegner llama nuestra atención sobre aquellos momentos en que tenemos un pensamiento perfectamente consciente pero que no sabemos de dónde viene, algo no tan infrecuente entre aquellos que acostumbramos a «tener la mente ausente»; según la acertada expresión de Wegner, tal pensamiento es *consciente pero no accesible* (Wegner, 2002, pág. 163). (¿Y ahora qué hago de pie en la cocina frente al armario? Sé que estoy en el sitio donde quería estar, pero ¿qué había venido a buscar?) En un momento así, *yo* he perdido de vista el contexto y, por lo tanto, la *raison d'être*, de este pensamiento, de esta experiencia consciente, motivo por el cual su significado (y eso es lo más importante) ha pasado a ser tan poco accesible para *mí* —el yo extendido que toma las decisiones— como lo pueda ser para cualquier tercera persona, cualquier observador «externo». En realidad, tal vez alguien que estuviera mirando me podría recordar* lo que me disponía a hacer. La posibilidad de que me recuerden lo que pensaba es crucial, puesto que eso es lo único que podría convencerme de que dicho observador tenía razón, de que eso era precisamente lo que *yo* estaba haciendo. Si el pensamiento o el proyecto es de alguien, es mío: me pertenece a mí porque le di origen y generé el contexto donde dicho pensamiento tiene sentido; ocurre simplemente que la parte de mí que está desconcertada es incapaz por el momento de tener acceso a la otra parte de mí que ha creado el pensamiento.

Podría decir, a modo de disculpa, que no era *yo mismo* cuando cometí ese error, o que olvidé lo que quería hacer, aunque no estamos sin duda

*** En inglés *remind* (literalmente algo parecido a «devolver la mente»), lo cual recoge la idea anterior de «tener la mente ausente» *(absent-minded).* (N. *del t.)***

ante la severa alteración del autocontrol que se observa en la esquizofrenia, en cuyo caso el paciente interpreta los propios pensamientos como voces ajenas. Esto no es más que una pérdida momentánea de contacto que interfiere en un plan perfectamente válido. Buena parte de lo que *yo* soy, buena parte de lo que hago y de lo que sé, procede de estructuras situadas abajo en la sala de máquinas, que son las que ponen en marcha las acciones. Si un pensamiento nuestro fuera *únicamente* consciente, y no fuera accesible *para esa maquinaria* (al menos para parte de ella, la parte que la necesita), entonces no podríamos hacer nada con él y nos quedaríamos repitiéndonos la maldita frase una y otra vez para nosotros mismos, para nuestro yo aislado. La conciencia aislada no puede hacer nada por sí misma. Y tampoco puede ser responsable de nada.

Tal como observa Wegner: «El hecho de que la gente tienda a olvidar ciertas tareas por el simple motivo de que ya las ha realizado señala una *pérdida de contacto* [la cursiva es mía] con sus intenciones iniciales una vez que las acciones han sido completadas, y, por lo tanto, una susceptibilidad de revisar las propias intenciones» (Wegner, 2002, pág. 167). ¿Una pérdida de contacto entre qué y qué? ¿Entre un yo cartesiano que «no hace nada» y un cerebro que toma todas las decisiones? No. Una pérdida de contacto entre la parte de nosotros que estaba al mando entonces y la parte de nosotros que está al mando ahora. Una *persona* tiene que ser capaz de mantenerse en contacto con las intenciones pasadas y las intenciones previstas, y una de las funciones principales de la ilusión de sí mismo que tiene el usuario del cerebro, a la que llamo «yo como centro de gravedad narrativa», es ofrecerme a *mí* un modo de interactuar conmigo mismo en otros momentos. Tal como lo expresa Wegner: «La voluntad consciente es particularmente útil, por lo tanto, como guía para orientarnos en nosotros mismos» (pág. 328). El cambio de perspectiva que debemos realizar para escapar de las garras del Teatro Cartesiano es hacernos a la idea de que el *yo,* ese yo extendido tanto espacial como temporalmente, puede controlar en cierta medida lo que ocurre de este lado de la barrera de la simplificación, en el lugar donde se toman las decisiones, y ése es el motivo por el que Wegner dice: «Ilusoria o no, la voluntad consciente es la guía de la persona hacia su responsabilidad moral en la acción» (pág. 341).

Sé que a muchas personas les resulta difícil comprender esta idea o tomársela en serio. Les parece algo así como un juego de espejos, un truco de magia verbal que escamotea la conciencia y el verdadero yo justo cuando iba a aparecer en escena. A muchas personas les parece, en la línea de Robert Wright, que este punto de vista niega la existencia de la conciencia

en lugar de explicar realmente su origen. ¿Dónde está la conciencia en este modelo? Pero ahí está, inadvertida en medio de todas las actividades que he descrito. Los contenidos mentales se hacen conscientes no porque entren en alguna especie de cámara especial en el cerebro ni porque sean traspasados a algún medio privilegiado y misterioso, sino cuando ganan las competiciones con otros contenidos mentales por el control del comportamiento, y en consecuencia por la generación de efectos duraderos (o, tal como se dice, de manera engañosa, por «entrar en la memoria»). Siendo como somos seres parlantes, y puesto que hablar con nosotros mismos es una de las actividades que más influencia tiene sobre nosotros, una de las formas más efectivas para que un contenido mental aumente su capacidad de influencia —aunque no la única— es hacerse con la parte lingüística de los controles. Todo esto tiene que suceder en el cerebro, en el «procesador central», pero no es necesario que ninguna instancia dirija su actividad. Tal como ha señalado Ainslie: «El ordenado mercado interno figurado por las teorías convencionales de la utilidad se convierte en una confusa pelea sin reglas» (Ainslie, 2001, pág. 40), el yo cartesiano se fragmenta en una insegura coalición, sin ningún rey ni juez que la presida.

> CONRAD: Supongamos que todos estos extraños procesos competitivos tienen lugar efectivamente en mi cerebro, y supongamos que, tal como dice usted, los procesos conscientes son simplemente los que ganan las competiciones. ¿Cómo hace eso que sean conscientes? ¿Qué ocurre con ellos para que yo sepa efectivamente de su existencia? Después de todo, lo que debemos explicar es mi conciencia, lo que yo sé desde el punto de vista de la primera persona.

Esta pregunta revela una profunda confusión, ya que presupone que *usted* es alguna *otra* cosa, alguna *res cogitans* cartesiana sobreañadida a todas las actividades del cuerpo y el cerebro. Usted, Conrad, no *es* otra cosa que esta organización de la actividad competitiva del cerebro entre una multitud de competencias que su cuerpo ha desarrollado. Usted conoce «automáticamente» estas cosas que ocurren en su cuerpo, porque si no lo hiciera, no sería su cuerpo.

Los actos y los hechos de los que puede hablarnos usted, y las razones que hay detrás de ellos, son suyas porque usted las ha creado... y porque ellas le han creado a usted. Usted no es otra cosa que aquel agente sobre cuya vida puede hablar. Puede hablar de ella con nosotros, y también con usted mismo. El proceso de autodescripción comienza en la infancia más

temprana, e incluye una buena dosis de fantasía desde el comienzo. (Piense en el personaje de los *Peanuts,* Snoopy, sentado en su casa de perro y pensando: «Aquí viene el as de la Primera Guerra Mundial, volando hacia la batalla».) Y continúa así toda la vida. (Piense en el camarero de la discusión de Jean Paul Sartre sobre la «mala fe» en *El ser y la nada* [1943], preocupado por encontrar la manera de vivir de acuerdo con su autodescripción como camarero.) Eso es lo que hacemos. Eso es lo que somos.[5]

Las exigencias de la comunicación no sólo hacen necesaria la clase de autocontrol que genera la ilusión del Teatro Cartesiano. También abren la psicología humana a una rica variedad de desarrollos ulteriores. El hecho de que las complejidades primarias de nuestro entorno no sean sólo otros agentes —depredadores, presas, rivales o parejas potenciales—, sino otros agentes capaces de *comunicarse* —amigos o enemigos potenciales, conciudadanos potenciales— tiene todavía otras implicaciones para la evolución de la libertad humana, que desarrollaremos en los capítulos restantes.

Capítulo 8

¿Exactamente cuándo y dónde tomamos las decisiones? Cuando examinamos de cerca las decisiones conscientes de una persona, descubrimos que esta búsqueda de precisión espaciotemporal no lleva a ninguna parte, lo que genera la ilusión de un yo aislado e impotente. La forma de restaurar el poder del yo, y, por lo tanto, su capacidad para la responsabilidad moral, es reconocer que sus tareas están distribuidas en el cerebro tanto en el espacio como en el tiempo.

Capítulo 9

¿Cuáles son los prerrequisitos para la autonomía, y cómo fue posible que se dieran? Para ser agentes morales debemos ser capaces de actuar en función de unas razones que sean nuestras razones, pero en el mejor de los casos podemos considerarnos unos razonadores imperfectos. ¿Podemos ser lo bastante racionales como para conservar la noción de nosotros mismos como genuinos agentes morales? Y, si es así, ¿cómo hemos llegado a serlo?

5. Algunas partes de los tres párrafos precedentes están tomadas, con revisiones, de Dennett, 1997b.

NOTAS SOBRE FUENTES Y LECTURAS COMPLEMENTARIAS

El artículo más reciente de Libet sobre este tema aparece en un volumen inspirado en sus experimentos, *The Volitional Brain* (Libet y otros, 1999), que incluye artículos psicológicos, neurológicos, teológicos, filosóficos y simplemente raros. El libro no tiene parangón en cuanto a apertura de miras, como lo demuestra el hecho de que incluya como artículo final una mordaz crítica de sí mismo, «A Review of *The Volitional Brain*», donde Thomas Clark (1999) expone de forma incisiva pero justa los principales errores y muchas de las confusiones de los artículos que lo preceden. Los filósofos han escrito bastante sobre la contradicción pragmática implícita en proposiciones como «*p* y nadie debería creer que *p*». Ahora disponen de un ejemplo real y a gran escala de la misma contradicción pragmática. (En realidad, Stephen Stitch ya se les había adelantado: el primer capítulo de su *Deconstructing the Mind* [1996] se propone explícitamente refutar todos los capítulos que le siguen, que no son sino reediciones de otros artículos, algunos de ellos escritos en colaboración con estudiantes suyos de posgrado. Es un ejemplo de cambio público de opinión que desearía ver emulado por otros filósofos, aunque me pregunto si sus varios coautores estaban tan dispuestos como él a abandonar el barco. En todo caso no lo dicen.) Mis propios comentarios a Libet incluyen el capítulo 6, «Tiempo y experiencia», de *La conciencia explicada* (1991a); otro artículo algo más técnico, escrito en colaboración con Marcel Kinsbourne, «Time and the Observer: The Where and When of Consciousness in the Brain», en *Behavioral and Brain Sciences,* 1991 (véase también el comentario de Libet en el mismo libro); y mis contribuciones, incluido un debate con Libet, en el compendio del simposio de 1993 de la Fundación CIBA, *Experimental and Theoretical Studies of Consciousness,* especialmente las páginas 134-135. Véase también la versión de Libet en Libet, 1996.

La literatura filosófica sobre los nefandos neurocirujanos que implantan sistemas de control remoto en el cerebro de las personas tiene su origen en el artículo clásico de Harry Frankfurt, «Alternative Possibilities and Moral Responsibility» (1969). Véase Kane, 2001, y cabe destacar también entre los libros recientes *Responsibility and Control: A Theory of Moral Responsibility* (1998), de John Martin Fischer y Mark Ravizza.

Resulta particularmente interesante estudiar la permeable frontera entre el aprendizaje individual y el «instinto» transmitido genéticamente

desde la perspectiva del Efecto Baldwin, o lo que C. H. Waddington llamó la asimilación genética, tema que traté tanto en *La conciencia explicada* (Dennett, 1991a) como en *La peligrosa idea de Darwin* (Dennett, 1995). Un libro de próxima aparición, compilado por Bruce Weber y David Depew, recoge una nueva hornada de reflexiones a propósito del Efecto Baldwin, que incluyen un extenso alegato mío en defensa del Efecto Baldwin: «The Baldwin Effect: A Crane, not a Skyhook» (Dennett, 2002b). Otras ideas que aparecen en este capítulo están desarrolladas en *Tipos de mente* (Dennett, 1996a), «Learning and Labeling» (Dennett, 1993) y «Making Tools for Thinking» (Dennett, 2000a).

Capítulo 9

Auparse a la libertad

Es la cultura lo que nos ha permitido convertirnos en lo que Aristóteles resumió con su célebre fórmula: animales racionales. ¿Cómo? Al hacer posible, una vez más, una división del trabajo y una distribución de la responsabilidad que ha dado lugar a nuevos niveles de sofisticación en el diseño a lo largo de la historia evolutiva.

SOBRE CÓMO CAPTAMOS LAS RAZONES Y LAS HICIMOS PROPIAS

> *Somos criaturas que preguntan por qué, igual con las normas que en otros terrenos. Queremos que la moral no sea un conjunto ciego de tabúes, sino algo que tenga un sentido (o tal vez más de un sentido, pero entonces queremos descubrir qué relaciones pueden tener entre sí estos sentidos y cómo podemos reconciliarlos).*
>
> ALLAN GIBBARD, *Wise Choices, Apt Feelings*

La conciencia humana fue hecha para compartir las ideas. Eso significa que la interfaz de usuario humano fue creada por la evolución, tanto biológica como cultural, y que surgió en respuesta a una innovación en el comportamiento: la actividad de comunicar proyectos y creencias, y comparar las experiencias respectivas. Esta interconexión convirtió los cerebros en mentes, e hizo posible una distribución de la autoría que es la fuente no sólo de nuestra enorme superioridad tecnológica respecto al resto de la naturaleza, sino también de nuestra moral. El último paso requerido para completar mi versión naturalista de la libertad y la responsabilidad moral es explicar el proceso de I+D que nos ha dado a cada uno de nosotros una perspectiva sobre nosotros mismos, un lugar desde el que *asumir* la responsabilidad. El

nombre de este punto arquimédico es el yo. Éste es un rasgo propio de los seres humanos que nos distingue como potenciales agentes morales, y no es ninguna sorpresa que el lenguaje esté implicado en él. Lo que resulta más difícil de comprender es cómo el lenguaje pudo traer consigo, una vez instalado en el cerebro humano, la construcción de una nueva arquitectura cognitiva capaz de *crear* un nuevo tipo de conciencia... y la moral.

La cuestión es tanto histórica como de justificación. Si fuera una cuestión meramente histórica la respuesta podría ser del tipo: érase una vez, hace muchos años, que unos alienígenas del espacio vinieron a la Tierra y nos hicieron tragar a todos unas pastillas de moral; desde entonces les enseñamos moral a nuestros hijos. O tal vez una versión ligeramente más realista: un retrovirus diezmó a nuestros antepasados homínidos, y resultó que los escasos supervivientes habían desarrollado un *gen para la apreciación de la justicia.* U otra todavía más realista: los memes de la moral surgieron por accidente hace unas decenas de miles de años, y se extendieron como una epidemia entre la población humana a nivel mundial. Incluso aunque alguna de estas fabulosas historias fuera cierta, nos dejaría sin la mitad de la explicación que necesitamos: ¿qué pasa con la justificación?

Por fortuna, el razonamiento darwinista es ideal para explicar cosas que persiguen «un fin». Cualquier explicación en términos de selección natural presupone una respuesta —sea la que sea— a la pregunta de *Cui bono?* Sin embargo, todavía tenemos que buscar nuevas respuestas a la pregunta darwinista por el *Cui bono,* puesto que el fin de la moral no se restringe manifiestamente al «bien de la especie» o «la supervivencia de nuestros genes» o nada parecido. Deberá ser algo que surja en el curso de nuestra propia constitución como la clase de sujetos que somos.

Uno de los rasgos desconcertantes de los procesos evolutivos descritos en los capítulos precedentes es la ausencia de nada parecido a la comprensión en los agentes cuyas inclinaciones se ven configuradas por dichos procesos. Es posible que estos agentes (o mejor aún, sus genes) sean los beneficiarios de ciertos instintos amigables, ciertas disposiciones genéricas *a cooperar,* pero eso no representa nada para ellos. No tienen por qué darse cuenta de las razones que hay detrás de los principios que dominan sus vidas, no tienen por qué apreciar sus razones virtuales y, por lo tanto, tampoco representárselas. La evolución de nuestra capacidad para reconocer dichas razones y reflexionar sobre ellas, y convertirlas en razones enteramente distintas, fue otra transición crucial en la historia evolutiva, e igual que todas las demás, tuvo que operar a partir de cosas que habían evolucionado para servir a otros propósitos.

La idea básica es bien conocida desde hace siglos. Según explica David Hume, comenzamos por lo que llama los motivos naturales: el apetito sexual, el afecto por los niños, una benevolencia limitada, el interés y el resentimiento (una lista que cualquier psicólogo evolutivo del siglo **XXI** vería con buenos ojos). Dichas disposiciones tienen detrás unas razones que no son *nuestras* razones, aunque han creado el escenario que hace posible nuestra práctica de pedir y dar razones. Tal como dijo Hume en su *Tratado de la naturaleza humana-*. «Si la naturaleza no nos hubiera ayudado en este particular sería inútil que los políticos hablaran de que algo fuera *honroso* o *deshonroso,* digno de *elogio* o de *repulsa.* Dichas palabras serían perfectamente ininteligibles» (Hume, 1739, pág. 500). Desde el principio nos encontramos con que aprobamos ciertas prácticas y actitudes —como si fueran en cierto sentido «intrínsecamente» buenas—, unas prácticas y actitudes que fueron configuradas a lo largo de milenios sin la intervención de ningún diseñador previsor, sino a partir de sus propias *raisons d'être.* Es posible que algunos de nuestros antepasados percibieran al menos vagamente los beneficios de algunos de estos hábitos y prácticas tan arraigados, pero ni siquiera esto es un requisito inexcusable, puesto que hay (al menos) tres maneras de explicar por qué los diseños heredados salieron a cuenta desde la perspectiva de la replicación diferencial: 1) que nuestras motivaciones naturales sean adaptaciones directamente ventajosas para los individuos que las poseen (una selección a nivel del individuo, el caso más o menos estándar); 2) que existiera una estructura grupal lo bastante desarrollada en las poblaciones humanas como para crear unas condiciones bajo las cuales aquellos grupos que siguieran inconscientemente ciertas prácticas prosperaran a costa de grupos constituidos menos favorablemente (selección grupal); o 3) que los memes de las motivaciones hayan estado compitiendo por el limitado número de nichos disponibles en los cerebros humanos e, igual que muchos otros de nuestros simbiontes, hayan terminado fijándose por alguna razón como rasgos estables de la ecología cultural humana. Todas ellas son formas «naturales», en el sentido de Hume, de justificar que estemos imbuidos de las motivaciones que sirven de fundamento para la siguiente ola de I+D, la ingeniería social deliberada, que sólo tiene unos pocos milenios de historia. Los motivos naturales, sostenía Hume, tienen «descendencia», la cual consiste en lo que llamó las virtudes «artificiales» de la moral (como por ejemplo la justicia). Hume veía la ética como una especie de tecnología humana, y veía la reflexión como el instrumento que nos proporciona la naturaleza para que podamos revisar nuestros instin-

tos naturales, potenciarlos con nuevos complementos cuyas razones (virtuales hasta que Hume y otros lograron captarlas y representarlas) tienen por objetivo aumentar todavía más nuestra libertad, sin poner en peligro nuestra seguridad. Unas gafas para el alma, podría decirse. Pero antes de pasar a esta nueva clase de I+D, deberíamos considerar a grandes rasgos cuál pudo ser el proceso evolutivo que hizo posible la transición de los agentes inconscientes a los agentes reflexivos dotados de mente.

Comencemos con la elegante «fábula evolutiva» que propone Brian Skyrms en su libro *Evolution of the Social Contract* (1996, pág. 3 y sigs.) sobre el juego de cómo repartir un pastel. Supongamos que usted y yo nos encontramos con un pastel de chocolate y nos lo queremos repartir. En lugar de pelearnos por él (una opción peligrosa para ambos) nos ponemos de acuerdo para resolver la cuestión con un simple juego: «Cada uno escribe cuál es su pretensión última sobre el pastel en una hoja de papel, la dobla y se la entrega a un árbitro. Si ambas suman más del cien por cien el árbitro se come el pastel. En cualquier otro caso cada uno obtiene lo que ha reclamado para sí. (Podemos suponer que si la suma es inferior al cien por cien el árbitro se queda con la diferencia)» (pág. 4). Tal como observa Skyrms, casi todo el mundo escogería el 50 %, la porción equitativa. (El árbitro no forma parte en realidad del modelo, es sólo una forma de completar el cuadro.) Y, en efecto, la teoría de juegos evolutiva muestra que la repartición del 50 por 50 es una *estrategia evolutivamente estable*, o EEE. «Una división equitativa será estable en cualquier dinámica que tienda a incrementar la proporción (o la probabilidad) de las estrategias con mayores beneficios porque cualquier desviación unilateral de una división equitativa supone menos beneficios» (pág. 11). Pero ésta no es la única EEE posible, observa Skyrms; hay muchas otras. Esto es lo que se conoce como el problema de las *trampas polimórficas:*

> Supongamos por ejemplo que la mitad de la población reclama $^2/_3$ del pastel y la mitad de la población reclama V_3. Llamemos a la primera estrategia *codiciosa* y a la segunda *modesta.* Un individuo codicioso tiene iguales probabilidades de encontrarse con un individuo codicioso o con uno modesto.* Si se encuentra con otro individuo codicioso no consigue nada porque sus pretensiones exceden del conjunto del pastel, pero si se encuentra con un individuo modesto, obtiene $^2/_¡$. Sus beneficios medios son V_3. Un individuo modesto, por otro lado, obtiene unos beneficios de $'/_3$ con independencia de con quién se encuentre.

*** Puesto que no hemos introducido aún ninguna correlación. *(N. del a.)***

Comprobemos si este polimorfismo constituye un equilibrio estable. En primer lugar, nótese que si la proporción de codiciosos aumentara, los codiciosos se encontrarían unos con otros con mayor frecuencia, y los beneficios medios para los codiciosos caerían por debajo del V_3 que tienen garantizado los modestos. Y si la proporción de los codiciosos disminuyera, los codiciosos se encontrarían más a menudo con modestos, con lo que los beneficios medios de los codiciosos aumentarían por encima de $^1/_3$. La retroalimentación negativa mantendrá las proporciones poblacionales de codiciosos y modestos en una igualdad. Pero ¿qué ocurriría si hubiera una invasión de estrategias mutantes? Supongamos que surge en la población un mutante *supercodicioso* que reclama más de $^2/_3$. Dicho mutante obtiene unos beneficios de 0 y se extingue. Supongamos que surge en la población un mutante *supermodesto* que reclama menos de V_3. Dicho mutante obtendrá lo que pide, que es menos de lo que obtienen el codicioso y el modesto, por lo que también se extinguirá (aunque más lentamente que el supercodicioso). La posibilidad que nos queda es que surja un mutante a medio camino entre los dos, que reclame más que el modesto pero menos que el codicioso. Reviste un interés especial el caso del mutante *equitativo* que reclama exactamente V_2. Todos estos mutantes se quedarían sin nada cuando se encontraran con un codicioso, y obtendrían menos que el codicioso al encontrarse con un modesto. En consecuencia obtendrían unos beneficios medios de menos de V_3 y todos ellos —incluido nuestro mutante equitativo— se extinguirían. El polimorfismo tiene fuertes propensiones a la estabilidad.

Ello es una mala noticia, tanto para dicha población como para la evolución de la justicia, puesto que nuestro polimorfismo es ineficiente. Aquí todo el mundo obtiene, de media, $^1/_3$ del pastel, mientras que otro tercio se echa a perder en los encuentros entre los codiciosos (Skyrms, 1996, págs. 12-13).

Skyrms observa también que en cuanto añadimos alguna correlación positiva al modelo, de modo que los representantes de cada tipo de estrategia tiendan a interactuar con los de su propia clase en lugar de emparejarse al azar, estos lamentables polimorfismos se vuelven menos atractivos (y por lo tanto más evitables). No es relevante qué aspectos del mundo sean los que contribuyen a potenciar esta correlación, pero los *agentes con mentes y cultura* están particularmente bien dotados para lograr este efecto, tal como demuestra Don Ross en una imaginativa «Historia de así fue» basada en el modelo de Skyrms:

Imaginemos una población que ha optado por una de las EEE polimórficas. El éxito continuado de los agentes codiciosos en este juego dependerá de su capacidad de convencer a los modestos para que eviten las interacciones con cualquier mutante equitativo que pudiera aparecer. Cabría esperar pues

de esta sociedad que desarrollara normas de justicia parecidas a las de Aristóteles. Dichas normas asociarían la «justicia» a la idea de que los modestos deberían respetar su posición natural y someterse a los codiciosos. Son normas muy familiares en numerosas sociedades humanas, tanto pasadas como presentes. Si dichos agentes son incapaces de realizar cálculos moderadamente sofisticados, o derivar con el pensamiento las implicaciones de los mismos, la población se quedará en dicho estadio. Se encuentra, después de todo, en un equilibrio EEE. Pero si dichos agentes son capaces de desarrollar algo de economía y comprender la lógica darwinista básica —no se requiere nada muy elaborado— se darán cuenta de que la EEE de todos equitativos es: a) más eficiente (tesis económica) y b) alcanzable de modo estable (tesis darwinista). No cuesta mucho imaginar lo que sucedería. Inicialmente, la mayor parte de la población vería la idea de la EEE de todos equitativos como una violación flagrante de la moral natural. Pero *unos pocos* de los modestos pasarían del reconocimiento de a) al concepto de su propia explotación. ¿Por qué no? Cualquier criatura con una cierta flexibilidad conceptual ensayaría dicho paso, aunque sólo fuera para desdecirse de la conclusión y someterse a la opinión pública. Algunos modestos que abrazaran dicha idea serían perseguidos; pero esto no haría sino contribuir aún más a la difusión del meme al dramatizar su importancia. Si los modestos ilustrados fueran capaces de reconocerse unos a otros, tendrían a su alcance una modalidad tranquila y efectiva de rebelión: sólo tendrían que jugar a la estrategia equitativa entre ellos, con lo que accederían a los beneficios superiores de este tipo de comercio. Después de todo, cuando hablamos de «mutantes equitativos» no tenemos por qué referirnos a bichos raros; cada vez que el meme equitativo se aloja en la mente de un modesto, tenemos un mutante. Permítasenos suponer que hasta el momento los mutantes sólo están movidos por el deseo de obtener más: todavía no han cuestionado *moralmente* las normas dominantes. Es posible que la belleza matemática de los resultados más eficientes resultara lo bastante atractiva para algunos modestos, e incluso para algunos codiciosos, como para que la persiguieran por sí misma. Esto vendría a complementar el interés personal y serviría para acelerar la dinámica, aunque no es estrictamente necesario.

La teoría de juegos evolutiva muestra que esta población evolucionará inevitablemente hacia la EEE de todos equitativos. Mucho antes de llegar a este punto, surgirá de manera *natural* el concepto de justicia como equidad, puesto que los equitativos promueven mejor su propio éxito si animan al ostracismo de los codiciosos. Inculcar el rechazo moral hacia las estrategias de los codiciosos será un paso natural —un buen truco muy evidente—, con sólo que estén biológicamente equipados para experimentar un rechazo *simple* hacia alguna cosa. Al final, la población verá su consenso anterior (si tienen la suficiente amplitud de perspectiva) como una especie de estadio infan-

til y amoral. Si no la tienen, decidirán que sus antepasados eran malos, y algunos individuos estúpidos e inseguros desaconsejarán la lectura de los libros supervivientes de la época anterior.

Veamos ahora qué es lo que ha ocurrido aquí. Dichos agentes experimentaron una evolución moral, mensurable usando un estándar objetivo. El primer paso necesario para ello fue alcanzar cierta noción de la lógica elemental darwinista. Ningún superhéroe moral, ningún Cristo ni ningún Nietzsche tuvo que venir a exhortarles para dar este paso. Bastó un poco de ciencia y de lógica. Al final del proceso, ¿saben algo estos agentes que sus antepasados desconocían? Sin duda: saben que la equidad es justa; *son* moralmente superiores a sus antepasados. Contra el principio de la Guillotina de Hume,* lo descubrieron gracias a ser difusores conscientes de memes capaces de pensar en términos hipotéticos, y gracias a que usaron algunas de esas capacidades para aprender algo de teoría evolutiva (Ross, correspondencia personal).

Por supuesto, no es necesario usar el lenguaje de los economistas profesionales para apreciar la tesis económica, y no es preciso ser explícitamente darwinista para ver cómo se puede pasar del primer estadio (una ineficiente trampa polimórfica) al segundo (una distribución equitativa) por una vía sostenible. Como siempre, bastará con una difusa versión medio comprendida, medio imaginada, para que encontremos gradualmente el camino de la inconciencia a la comprensión. El propio Darwin llamó nuestra atención hacia la importancia de lo que llamaba *selección inconsciente*, como paso intermedio entre la *selección natural* y lo que llamó la *selección metódica:* la deliberada, previsora e intencionada «mejora de la raza» por parte de los criadores de animales y plantas. Darwin observó que la línea que separa la selección inconsciente de la metódica era también una frontera borrosa y gradual:

> El primer hombre que seleccionó una paloma con una cola ligeramente más larga nunca soñó en qué terminarían convirtiéndose los descendientes de dicha paloma a través de una selección continuada, en parte inconsciente, en parte metódica (Darwin, 1859, pág. 39).

Y tanto la selección inconsciente como la metódica son sólo casos especiales de un proceso más amplio, la selección natural, dentro del cual el impacto de la inteligencia y de la capacidad de elección humanas es prácticamente nulo. Desde la perspectiva de la selección natural, los cambios en los linajes debidos a la selección inconsciente o metódica no son sino

*** El principio de que no se puede derivar el «deber ser» del «ser». (N.** *del a.)*

cambios que tienen lugar en un entorno donde la actividad humana es uno de los factores selectivos más destacados. Recientemente se ha unido un nuevo miembro a este catálogo de procesos de selección natural: la ingeniería genética. ¿En qué se diferencia de la selección metódica de la época de Darwin? Es menos dependiente de las variaciones preexistentes en el acervo genético y genera nuevos genomas de forma más directa, sin necesitar un proceso tan lento y directo de ensayo y error. Nuestra capacidad de predicción es cada vez mayor, pero, incluso en este estadio, si examinamos de cerca las prácticas de laboratorio, veremos que hay todavía una buena dosis de ensayo y error en la búsqueda de las mejores combinaciones de genes.

Podemos usar los tres niveles de selección genética de Darwin, más el cuarto nivel más reciente de la ingeniería genética, como modelo para definir otros cuatro niveles paralelos de selección *memética* en la cultura humana. Los primeros memes eran seleccionados de manera natural, y prepararon el camino para memes seleccionados inconscientemente —memes «domesticados» sin darnos cuenta, por decirlo así—, a los que siguieron los memes metódicamente seleccionados, en cuyo desarrollo desempeñaron un papel importante la previsión y los proyectos, aunque los mecanismos subyacentes eran sólo vagamente vislumbrados, y la mayor parte de la experimentación consistía en la búsqueda de variaciones simples de los temas existentes, hasta la situación actual, donde la ingeniería memética es una de las grandes empresas humanas: el proyecto de diseñar y difundir sistemas completos de cultura humana, teorías éticas, ideologías políticas, sistemas de justicia y de gobierno, un sinfín de diseños de vida social en competencia entre sí. La ingeniería memética es una innovación muy reciente en la historia de la evolución en este planeta, pero sigue siendo unos milenios más vieja que la ingeniería genética: algunos de sus primeros y más célebres productos fueron *ha república* de Platón y la *Política* de Aristóteles.

No somos simples criaturas popperianas, capaces de anticipar el futuro e imaginar alternativas posibles y sus resultados probables, sino criaturas gregorianas, capaces de usar además las herramientas de pensamiento que nuestras culturas nos inculcan durante la infancia y más adelante (Dennett, 1995, págs. 377 y sigs.). Nos enfrentamos a los dilemas humanos con la ayuda de un equipaje compartido de preceptos memorizados y que tenemos siempre en la punta de la lengua. Incluso los cuentos de hadas y las fábulas de Esopo tienen una función valiosa a la hora de canalizar adecuadamente la atención de los niños. Una de las ra-

zones por las que raramente nos metemos en un callejón sin salida o segamos la hierba bajo nuestros pies es que hemos escuchado algún cuento divertido o memorable acerca de un tipo que hizo precisamente eso. Y si seguimos la regla de oro, o los Diez Mandamientos, no hacemos sino reforzar nuestros instintos naturales subyacentes con complementos que tienden a favorecer ciertas conceptualizaciones de las situaciones a las que nos enfrentamos. Pero buena parte de esta tradición ha evolucionado sin ningún autor explícito, y se ha transmitido sin ninguna apreciación explícita de su utilidad, hasta un momento relativamente reciente.

LA INGENIERÍA PSÍQUICA Y LA CARRERA ARMAMENTISTA DE LA RACIONALIDAD

> *Lo que hice fue adoptar el punto de vista de un ingeniero psíquico que hubiera recibido el encargo de diseñar nuestras normas para obtener unas ventajas que pudieran ser reconocidas por todos.*
>
> ALLAN GIBBARD, *Wise Choices, Apt Feelings*

Una vez que hemos captado las razones virtuales que hay detrás de las motivaciones naturales, y hemos desarrollado una representación de las mismas que añadir a las demás representaciones de todos los artificios que soñamos en el curso de nuestras reflexiones, ya no nos vemos limitados por el ineficiente, derrochador e inconsciente proceso de la selección natural. Podemos aspirar a sustituir un equilibrio basado en la pura capacidad replicativa por un *equilibrio reflexivo* propio de agentes racionales implicados en una actividad colectiva de persuasión mutua. Este paso de un proceso ciego de ensayo y error a un sistema de (re)diseño inteligente es, tal como he sugerido, una transición importante en la historia evolutiva, que abre nuevas dimensiones de oportunidades antes literalmente inimaginables, para bien o para mal. Hasta el nacimiento de la ética, la I+D darwinista había avanzado durante miles de millones de años sin la menor previsión, en su lento ascenso por la pendiente del Monte Improbable (Dawkins, 1996). Cada vez que los linajes llegaban a alguna cima local en el paisaje de la adaptación, sus miembros no tenían ninguna forma de preguntarse si habría o no mejores cimas a uno u otro lado del valle. En el marco de sus paisajes meramente físicos, los individuos más audaces y

perceptivos podían hacer algo equivalente a diseñar el objetivo de llegar al otro lado del río, o hasta esa franja visible de hierba comestible en aquella colina de allá, pero hubo que esperar a nuestra llegada para que encontraran expresión preguntas más remotas sobre el sentido de la vida y sobre la mejor manera de alcanzarlo. Somos la única especie cuyos miembros son capaces de ir más allá del paisaje físico e *imaginar* el paisaje adaptativo de las posibilidades, los únicos capaces de «ver» más allá de los valles hacia otras cimas concebibles. El mero hecho de que hagamos lo que estamos haciendo —tratar de esclarecer si nuestras aspiraciones éticas tienen algún anclaje sólido en el mundo que la ciencia está desvelando— demuestra hasta qué punto somos distintos de las demás especies.

Podemos concebir (al menos eso nos parece) mundos mejores y aspirar a alcanzarlos. ¿Estamos seguros de que esos mundos serían mejores? ¿En qué sentido? ¿Según qué criterios? Según los nuestros. Nuestra capacidad de reflexionar nos ofrece —sólo a nosotros— tanto la oportunidad como la competencia necesaria para evaluar los fines, no sólo los medios. Debemos usar nuestros valores actuales como puntos de partida para cualquier posible reevaluación de nuestros valores, pero la perspectiva que nos ofrece la cima actualmente alcanzada ya nos permite formular, criticar, revisar y —si tenemos suerte— acordar una serie de principios de diseño para la vida en sociedad. Podemos contemplar tentadoras cimas utópicas harto distintas de todo cuanto conocemos actualmente. ¿Podemos llegar hasta alguna de ellas? ¿Estamos seguros de querer intentarlo? Sería una lástima si al final no consiguiéramos alcanzarlas, pero no sería en ningún caso una ofensa a la razón. Uno de los problemas de diseño más difíciles que tenemos delante es descubrir y aislar los factores determinantes en el terreno de la política, el arte de lo posible. Podría darse el caso lamentable de que estuviéramos atrapados en el mejor de los mundos posibles, dadas las circunstancias históricas, pero también en este caso podríamos introducir alguna corrección en nuestro diseño actual que nos diera alguna esperanza de alcanzar cimas más altas. Y a diferencia de lo que ocurre con las demás especies, todo esto son problemas que se nos plantean *a nosotros*. Trabajamos en ellos, les dedicamos tiempo y energía. Reunimos información relevante para resolverlos, exploramos variaciones de los mismos, y debatimos sus méritos respectivos, conscientes de que nuestras reflexiones contribuirán efectivamente a determinar la trayectoria que tomará nuestro futuro.

Esto ofrece, finalmente, un marco naturalista en el que tienen sentido las cuestiones tradicionales sobre moral. Nuestro viaje evolutivo nos ha

traído hasta el terreno tradicional del debate y la investigación filosófica y política, donde muchas ideas distintas compiten por ganarse nuestra aprobación. La ética es un campo vasto y complejo, una competición que no pretendo arbitrar o ni siquiera ampliar con ninguna aportación en este libro, más allá de unas pocas sugerencias sobre algunos rastros fósiles del viaje que todavía distorsionan el pensamiento ético. Una de nuestras tareas más urgentes, como ingenieros psíquicos, es ver si podemos fundamentar el concepto de agente moralmente responsable, un agente que, a diferencia del cooperativo perro de las praderas, del fiel lobo o del amable delfín, escoge libremente en función de razones meditadas y que puede ser considerado responsable de sus elecciones. Hemos esbozado el desarrollo evolutivo de las pautas que constituyen el entorno conceptual que hace posible el surgimiento de un concepto de este tipo —el aire que respiramos—, pero debemos examinar con más detalle cómo puede elevarse un individuo a una posición tan elevada. ¿Hay alguien realmente cualificado para ello? ¿No nos está enseñando la psicología que estamos muy lejos de ser los agentes racionales que pretendemos ser?

Alien Funt fue uno de los grandes psicólogos del siglo XX. Sus experimentos y demostraciones informales en *Candiel Camera* nos enseñaron tanto sobre la psicología humana y sus sorprendentes limitaciones como pudiera hacerlo la obra de cualquier psicólogo académico. Aquí tenemos uno de sus mejores experimentos (según lo recuerdo años más tarde): Funt montó un puesto especial en un lugar céntrico de unos grandes almacenes y lo llenó de mangos nuevos y relucientes de carrito de golf. Eran unos tubos fuertes y brillantes de acero inoxidable, de unos sesenta centímetros de largo, con una ligera curvatura en el centro, con rosca en un extremo (para atornillarlo en el lugar correspondiente de nuestro carrito) y con un sólido pomo esférico de plástico en el otro. En otras palabras, la pieza de acero inoxidable más inútil que se pueda imaginar, a menos que uno tuviera un carrito de golf al que le faltara el mango. Luego puso un cartel. No identificaba el contenido de la parada, sino que simplemente decía: «Rebaja del 50 %. ¡Sólo hoy! 5,95 dólares». Algunas personas los compraron, y cuando les preguntaban por qué lo habían hecho improvisaban cualquier explicación. No tenían la menor idea de qué era aquello, pero era un objeto bonito, y ¡toda una ganga! Esa gente no estaba bebida ni tenía ninguna lesión cerebral; eran adultos normales, nuestros vecinos, nosotros mismos.

Cuando nos asomamos al abismo que abre ante nosotros una demostración de este tipo nos entra una risa nerviosa. Tal vez seamos listos, pero

ninguno de nosotros es perfecto, y aunque tal vez usted o yo no cayéramos en el viejo truco del mango del carrito de golf, sabemos que hay otras variantes de este truco en las que hemos caído y en las que sin duda volveremos a caer en el futuro. Cuando descubrimos lo imperfecta que es nuestra racionalidad, nuestra propensión a dejarnos arrastrar en el espacio de las razones por cosas que no son razones conscientemente apreciadas, nos asalta el miedo de que tal vez no seamos libres después de todo. Tal vez nos estemos engañando. Es posible que nuestra aproximación a una perfecta facultad kantiana de la razón práctica se quede tan corta que nuestra orgullosa autoproclamación como agentes morales no sea sino un delirio de grandeza.

Nuestros fallos en estos casos son sin duda fracasos en nuestra aspiración a la libertad, a responder como nosotros querríamos ante las oportunidades y las crisis que nos presenta la vida. Por esta razón son tan incómodos, porque ésta es una de las formas valiosas y deseables de la libertad. Nótese que la demostración de Funt no nos impresionaría en lo más mínimo si sus sujetos no fueran personas, sino animales: perros, lobos, delfines o monos. Apenas puede sorprendernos que se pueda engañar a una simple bestia para que opte por algo brillante y vistoso, pero que no es lo que realmente quiere, o lo que debería querer. Partimos de la base de que los animales «inferiores» viven en el mundo de las apariencias, movidos por «instintos» y capacidades perceptivas que poseen una gran eficiencia dentro de su contexto, pero que quedan fácilmente en evidencia en circunstancias inusuales. Nosotros aspiramos a un ideal superior.

A medida que aprendemos cosas sobre las debilidades humanas y sobre la forma en que pueden aprovecharse de ellas las tecnologías de la persuasión, parece como si nuestra tan cacareada autonomía no fuera sino un mito insostenible. «Tome una carta cualquiera», dice el mago, y consigue hábilmente que uno tome la carta que él ha escogido previamente. Los vendedores conocen mil formas de hacer que nos decidamos a comprar ese coche, ese vestido. Según parece, bajar la voz funciona extraordinariamente bien: «Para mi gusto le queda mejor el verde». (Tal vez le interese recordarlo la próxima vez que un vendedor le susurre algo.) Nótese que se trata de una carrera de armamentos donde cada estrategia se ve contestada con una contraestrategia. Yo no he hecho más que reducir un poco la efectividad del truco del susurro para aquellos que recuerden mi consejo. Resulta fácil descubrir el ideal de racionalidad que está detrás de esta batalla: *caveat emptor,* decimos, el riesgo es del comprador. Se trata

de una política que presupone que el comprador es lo bastante racional como para no dejarse engañar por las lisonjas del vendedor, pero como ya no tenemos una fe inocente en este mito suscribimos una política basada en el *consentimiento informado,* que prescribe para la validez del acuerdo la presentación explícita de todas las condiciones relevantes en un lenguaje claro y comprensible. Pero también reconocemos que dicha estrategia es fácilmente eludible —la estratagema de la letra pequeña, la pomposa e impresionante jerga especializada—, por lo que podríamos prescribir todavía más condiciones para servirle la información a cucharadas al indefenso cliente. ¿En qué punto habremos superado el mito del «consentimiento adulto» en nuestra «infantilización» de la ciudadanía? Cuando oímos hablar de propuestas para adaptar los mensajes a determinados grupos o individuos mediante la inclusión de imágenes, explicaciones, facilidades y advertencias específicas, tal vez estemos tentados de condenarlo como paternalismo y considerarlo igualmente subversivo para el ideal de libertad según el cual somos agentes racionales kantianos, responsables de nuestro propio destino. Pero al mismo tiempo deberíamos reconocer que el entorno donde vivimos no ha cesado de actualizarse desde el origen de la civilización, a través de una cuidadosa preparación y de la instalación de toda clase de señales y advertencias en nuestro camino, para ayudarnos y hacernos las cosas más fáciles a nosotros, imperfectos electores. Aprovechamos sin dudarlo las mejoras que nos sirven *a nosotros* —eso es lo bueno de la vida civilizada—, pero acostumbramos a criticar las que necesitan los demás. En cuanto comprendemos que esto es una carrera de armamentos, resulta más fácil abandonar el absolutismo que sólo ve dos posibilidades: o somos perfectamente racionales o no somos racionales en absoluto. Dicho absolutismo promueve el miedo paranoico a que la ciencia podría estar a punto de demostrar que nuestra racionalidad es una ilusión, por muy benigna que pueda resultar desde ciertas perspectivas. Dicho miedo, a su vez, otorga un atractivo injustificado a cualquier doctrina que prometa mantener la ciencia a raya, conservar nuestras mentes como un espacio misterioso y sacrosanto. En realidad, somos extraordinariamente racionales. Somos lo bastante racionales, por ejemplo, como para ser muy buenos en el diseño de estrategias para ponernos trampas unos a otros, para buscar fisuras cada vez más sutiles en nuestras defensas racionales, un juego del escondite que no conoce ninguna pausa ni límite temporal.

Pero ¿cómo podemos conseguir ser lo bastante buenos en esto? Para dar una buena respuesta a esta pregunta es preciso defenderse de toda

clase de paradojas (Súber, 1992). Si somos libres, ¿somos responsables de ser libres, o es sólo cuestión de suerte? Tal como vimos en el capítulo 7, los cooperadores capaces de resolver los problemas de compromiso y crearse una reputación como agentes morales disfrutan de los muchos beneficios de ser miembros respetados de la sociedad, pero si uno no ha llegado todavía a este estatus, ¿qué esperanza le queda? ¿Qué deberíamos sentir hacia los frecuentes traidores que hay entre nosotros: desprecio o compasión? Las fronteras creadas por los procesos evolutivos tienden a ser porosas y graduales, con casos intermedios que cubren los pasos entre una categoría y la siguiente, pero no podemos emular a la Madre Naturaleza en su rechazo a introducir categorizaciones. Nuestros sistemas morales y políticos parecen obligarnos a clasificar a las personas en dos categorías: aquellos que son moralmente responsables y aquellos que están excusados de ello por no estar cualificados. Sólo los primeros son candidatos válidos para el castigo, para que se les pidan responsabilidades por sus acciones. ¿Cómo podemos decidir dónde marcar la línea? Ciertas acciones estúpidas y ciertos hábitos y rasgos de carácter que descubrimos en nosotros mismos pueden hacernos dudar de si una categorización como ésta puede ser algo más que un mito conveniente, algo así como el odioso mito de los metales de Platón, una estratagema pública pionera para mantener la paz en su República. Algunas personas han nacido de Oro, mientras que otras deberán contentarse con ser de Plata o de Bronce. Podría parecer, por ejemplo, que la teoría política adopta la estrategia de mantener una cierta proporción de castigos dentro de la sociedad para hacer creíbles las prohibiciones que mantienen controlados (en cierta medida) a los agentes racionales, una estrategia que estaría sin embargo condenada a la hipocresía. Aquellos que terminan siendo castigados pagan doblemente puesto que no eran realmente responsables de las acciones que, según nuestras beatas condenas, habrían cometido por su libre voluntad, y no son sino cabezas de turco a quienes la sociedad inflige un daño deliberado para sentar un vivido ejemplo ante los demás ciudadanos mejor dotados para el autocontrol. ¿Qué condiciones debe cumplir una persona para que podamos considerarla genuinamente culpable de sus fechorías? Y ¿hay alguien que las cumpla realmente?

CON ALGO DE AYUDA DE MIS AMIGOS

> *Las cosas que promete el romanticismo están muy lejos de ser ciertas: y, sin embargo, sólo por creerse una criatura un poco inferior al querubín el hombre se ha convertido, por una interminable sucesión de pequeños avances, en un ser netamente superior al chimpancé.*
>
> JAMES BRANCH CABELL, *Beyond Life*

> *Haz ver que lo consigues hasta que lo consigas.*
>
> **Eslogan de Alcohólicos Anónimos**

En el capítulo 4 consideramos y rechazamos la propuesta de Robert Kane de detener la amenaza de una regresión al infinito mediante el recurso a ciertos momentos más bien mágicos —las acciones autoformativas, o AA—, unos puntos en los que termina la cadena y el universo contiene el aliento mientras una indeterminación cuántica le permite «hacerlo usted mismo», crearse a sí mismo como agente moralmente responsable (y podría haber hecho otra cosa). La solución de Kane no funciona porque no podemos detener la regresión mediante la invocación de un mamífero primordial, mediante la invención de una diferencia especial que es «esencial» y sin embargo invisible. Una persona que decide desde una genuina aleatoriedad cuántica y su gemela que lo hace desde una pseudoaleatoriedad no difieren en ningún aspecto discernible que pudiera suponer una diferencia tan especial. Tal como revela un examen atento de la cuestión, nunca podríamos saber si habíamos logrado experimentar una genuina AA, por lo que su relevancia moral se disuelve, y persiste la amenaza de la regresión. ¿Cómo es posible pues que hayamos llegado a la agencia moral partiendo de la inconsciencia amoral de un niño, si no es por un milagroso salto de autocreación? Nadie se sorprenderá de que mi respuesta invoque tópicos darwinistas relacionados con el azar, el entorno y el gradualismo. Con un poco de suerte, y también un poco de ayuda de nuestros amigos, podremos poner a trabajar nuestro considerable talento innato y auparnos a la agencia moral, centímetro a centímetro.

El proceso básico quedó esbozado ya en el capítulo 8: un yo humano propiamente dicho es una creación en buena medida inconsciente de un proceso de diseño interpersonal por el que animamos a los niños a comunicarse y, en particular, a incorporarse a nuestra práctica de pedir y dar razones, y luego a razonar lo que hacen y por qué lo hacen. Para que esto

funcione, es necesario partir de las materias primas adecuadas. No lo conseguiremos si lo probamos con nuestro perro, por ejemplo, ni siquiera con un chimpancé, tal como lo han demostrado numerosos proyectos obstinados y entusiastas a lo largo de los años. Algunos niños humanos tampoco están a la altura de lo que se les pide. El primer umbral en el camino hacia la personalidad es simplemente la cuestión de si los propios cuidadores consiguen criar a un comunicador. Aquellos a quienes no se les encienden las luces de la razón, por un motivo u otro, quedan relegados a un estatus indiscutiblemente inferior. No es culpa suya, es simplemente mala suerte. Pero ya que entramos en el tema de la suerte, tratemos de afinar nuestros criterios. Todo ser vivo es, desde una perspectiva cósmica, increíblemente afortunado por estar vivo. La mayoría de los seres vivos que han vivido, el 90 % e incluso más, han muerto sin dejar ninguna descendencia viable, pero ni uno solo de nuestros antepasados, que se remontan hasta el origen de la vida en la Tierra, padeció este infortunio tan frecuente. Somos hijos de una cadena ininterrumpida de ganadores que se remonta a miles de millones de generaciones, y cuyos integrantes fueron, en cada generación, los más afortunados entre los afortunados, uno entre cientos o miles o incluso millones. De modo que por mucha mala suerte que podamos tener en alguna ocasión, nuestra mera presencia en el planeta testifica el papel que ha desempeñado la suerte en nuestro pasado.

Superado el primer umbral, las personas revelan una amplia diversidad de talentos ulteriores, tanto para pensar como para hablar o ejercer autocontrol. Parte de esta diferencia es «genética» —debida principalmente a diferencias en el conjunto particular de genes que componen sus genomas—, parte es congènita pero no directamente genética (debida a la malnutrición o a la adicción de la madre a las drogas, o al síndrome alcohólico fetal, por ejemplo) y parte no tiene ninguna causa en absoluto, tal como descubrimos en el capítulo 3: es el resultado del azar. Ninguna de estas diferencias en nuestra herencia está bajo nuestro control, por supuesto, puesto que ya estaban allí antes de nuestro nacimiento. Y es cierto que los efectos previsibles de algunas de ellas son inevitables, pero no todos (y cada año que pasa son menos). Tampoco es en ningún sentido culpa nuestra que naciéramos en un entorno determinado, rico o pobre, que fuéramos niños mimados o maltratados, que estuviéramos en la primera línea de la salida o en la última. Y todas esas diferencias, que son muchas, también tienen efectos diversos: algunos son evitables y otros inevitables, algunos dejan cicatrices para toda la vida y otros tienen un efecto transitorio. Muchas de las diferencias que sobreviven tienen, en cualquier caso,

una importancia desdeñable en comparación con lo que nos importa aquí: el segundo umbral, el de la responsabilidad moral (a diferencia de lo que sucede, por ejemplo, con el genio artístico). No todo el mundo puede ser un Shakespeare o un Bach, pero casi todo el mundo puede aprender a leer y escribir lo suficientemente bien como para convertirse en un ciudadano informado.

Cuando W. T. Greenough y F. R. Volkmar (1972) demostraron por primera vez que las ratas que se criaban en un entorno rico en juguetes, aparatos para hacer ejercicio y oportunidades para realizar exploraciones vigorosas tenían un mayor número de conexiones neurales y unos cerebros más grandes que las ratas criadas en un entorno vacío y restrictivo, algunos padres y educadores se apresuraron a llevar a la práctica este importante descubrimiento, y comenzaron a preocuparse por si el niño tenía suficientes juguetes en la cuna. En realidad, siempre hemos sabido que un niño que se haya criado solo en una habitación vacía y sin juguetes quedará seriamente afectado, y nadie ha demostrado aún que tener dos juguetes o veinte o doscientos suponga alguna diferencia perceptible a largo plazo en el desarrollo cerebral del niño. Sería extraordinariamente difícil de demostrar, a causa de las muchas influencias que intervienen en el proceso, algunas de ellas planificadas y otras fortuitas, que podrían hacer y deshacer cien veces al año el efecto que nos interesa a lo largo del desarrollo del niño. Deberíamos dedicar nuestros mejores esfuerzos a la investigación, pues es *posible* que algún otro factor desempeñe un papel más importante del que pensábamos (y, por lo tanto, sería un objetivo más apropiado para dirigir nuestros esfuerzos de evitación). Pero podemos estar bastante seguros de que la mayoría de estas diferencias en las condiciones de partida, si no todas, se desvanecen en la bruma estadística con el paso del tiempo. Igual que sucede con los lanzamientos de moneda, los resultados no tienen por qué revelar ninguna relación causal privilegiada. En cuanto hayamos conseguido aislar todos los factores hasta donde sea posible a través de un cuidadoso estudio científico, podremos decir con la debida certeza qué influencias pueden ser necesarias para compensar qué limitaciones, y sólo entonces estaremos en posición de realizar los juicios de valor que todo el mundo está tan ansioso por emitir.

Tom Wolfe, por ejemplo, deplora el uso de Ritalin (metilfenidato) y otras metanfetaminas para corregir el trastorno por déficit de atención con hiperactividad en los niños. Lo hace sin pararse a considerar la abundante evidencia científica que apoya la tesis de que algunos niños padecen un desequilibrio de dopamina en sus cerebros —fácilmente corregible (evita-

ble)— que genera un trastorno en el departamento de autocontrol con la misma certeza que lo hace la miopía en el departamento de la visión.

> Una generación entera de niños norteamericanos, desde las mejores escuelas privadas del noreste hasta las peores escuelas públicas basura de Los Angeles y San Diego, estaba ahora colgada del metilfenidato, que recibían diariamente de manos de su camello particular, la enfermera del colegio. ¡Norteamérica es un país maravilloso! ¡Lo digo en serio! ¡Ningún escritor honesto pondría en duda esta afirmación! ¡La comedia humana nunca se queda sin material! ¡Nunca deja que te aburras!
>
> Mientras tanto, la noción de un yo —un yo que ejerce la autodisciplina, pospone la satisfacción, contiene el apetito sexual, se refrena ante la agresión y el comportamiento criminal—, un yo que pueda hacerse más inteligente y auparse a las cimas de la vida por sus propios medios gracias al estudio, la práctica y la perseverancia a pesar de los grandes obstáculos en el camino, esta anticuada idea (¿qué significa auparse, por amor de Dios?) del éxito alcanzado gracias al esfuerzo y al auténtico valor se está perdiendo, perdiendo... perdiendo (Wolfe, 2000, pág. 104).

Este pasaje típicamente grandilocuente contiene cierta ironía poco frecuente en su autor. Me pregunto si Wolfe recomendaría un severo régimen de ejercicios oculares y cursos para «Aprender a vivir con la miopía» en lugar de darles unas gafas a los miopes. Wolfe termina por declamar la versión del siglo XXI del viejo lema: si Dios hubiera querido que voláramos, nos hubiera dado alas. Tan nervioso le pone el imaginario hombre del saco del determinismo genético que es incapaz de ver que el aupamiento que pretende proteger, la fuente misma de nuestra libertad, se ve potenciada antes que amenazada por una desmitologización del yo. El conocimiento científico es el camino real —el único camino— hacia la evitabilidad. Tal vez hayamos encontrado aquí el verdadero rostro del miedo secreto que se esconde detrás de algunos de los gritos de: *¡detengan a ese cuervo!* No es que la ciencia vaya a robarnos nuestra libertad, sino que va a darnos demasiada libertad. Si nuestro niño no tiene el mismo «auténtico valor» que el hijo del vecino, tal vez podamos comprarle algo de valor artificial. ¿Por qué no? Éste es un país libre, y la mejora personal es uno de nuestros mayores ideales. ¿Qué importancia habría de tener que hagamos todas nuestras mejoras a la manera antigua? Todas éstas son preguntas importantes, y sus respuestas no son evidentes. Pero deberían abordarse de frente y sin distorsiones motivadas por un injustificado afán de acallarlas.

En *Elbow Room* comparé las diferencias tanto genéticas como ambientales que se dan entre los participantes en la atropellada salida de una maratón, en la que algunos corredores comienzan varios metros por detrás de otros, aunque todos se dirigen a la misma línea de meta. Defendí que el sistema era justo, puesto que en una carrera tan larga «una ventaja inicial tan relativamente pequeña no tiene ninguna importancia, pues cabe esperar que otras influencias fortuitas puedan tener efectos aún mayores» (Dennett, 1984, pág. 95). Esto es cierto, pero subestima el papel de las influencias *no* fortuitas que se dan en la carrera hacia la responsabilidad adulta. Alcanzar el estatus de persona es un esfuerzo colectivo, donde el público y los entrenadores desempeñan un papel importante, al enriquecer el entorno con una especie de andamiaje diseñado (inconscientemente) para sacar lo mejor de nosotros. Más importante aún que el suministro de juguetes adecuados para el desarrollo, e incluso que una nutrición adecuada, es el conjunto de actitudes y prácticas que observa un niño en su entorno y en los que termina por participar. Hay una abundante evidencia científica que apoya la hipótesis de que los niños expuestos a personas violentas, mentirosas y rudas —tanto o más si son sus compañeros que si son sus padres— tienden a perpetuar dichos rasgos de carácter. También es importante considerar el lado bueno de la historia: aquellos que tienen la suerte de criarse en una sociedad libre, en compañía de personas razonables, sinceras y cariñosas, tienden a aspirar a esos mismos ideales. La crianza sí marca una gran diferencia.

Es un error reducir los efectos de la crianza a la «educación moral», como si la clave para garantizar que los propios pupilos se conviertan en adultos responsables fuera prestar la debida atención a un catecismo u otro. Disponer de un *vademecum* de preceptos condensados es algo sin duda útil, pero antes de eso se ha instalado ya un conjunto más potente de influencias, las cuales canalizan hasta nuestros más leves pensamientos. Somos conscientes a medias de que la mayor parte de lo que les decimos a nuestros hijos antes de que aprendan a hablar es como sí no lo hubiéramos dicho, pero no todo. Parte de ello se queda ahí. ¿Qué *quieres*? ¿Eso te da *miedo*? ¿Dónde te *duele? ¿Sabes* dónde está el conejito? ¿Quieres *tomarme el pelo*? «No te preocupes, ya crecerás», dice mamá, mientras le pone a su niño unas ropas heredadas de su hermano que le van grandes, y lo mismo puede decirse en buena medida de las disposiciones en cierto modo demasiado grandes que los adultos nos imponen cuando somos niños. Sin duda, crecemos hasta que nos van bien, y las hacemos nuestras, y pasamos a hacerlas nosotros, y nos convertimos de este modo en agen-

tes como los mayores. Cuanto más seriamente nos tomemos a nuestros hijos como participantes en la práctica de pedir y dar razones, tanto más seriamente acabarán por tomársela ellos.

Esta tendencia a *dar las cosas por supuestas*, a suponer más competencia en el diseño de nuestros jóvenes interlocutores de la que justificarían los fríos hechos, es un añadido extraordinariamente valioso al arsenal darwinista de trucos I+D. El hecho de que nosotros, los seres humanos, no seamos unos relojeros ciegos, sino unos educadores de personas dotados de una gran visión, capaces además de reflexionar sobre lo que vemos y de extraer inferencias sobre lo que queremos ver en el futuro, hace que seamos mucho más fáciles de rediseñar, primero por los demás y luego por nosotros mismos, que ningún otro organismo que haya evolucionado hasta ahora en el planeta. Consideremos por ejemplo el fenómeno de «sacar lo mejor de uno mismo». Más allá de cualquier instrucción que hayamos recibido en este sentido, sea formal o informal, casi siempre adaptamos nuestro comportamiento para que se ajuste a (lo que consideramos) las exigencias sociales del momento. Aparte de algunos raros espíritus libres que parecen genuinamente indiferentes a las presiones sociales, la gente debe realizar un gran esfuerzo de disciplina para frustrar deliberadamente las expectativas de aquellos que les rodean. Esta presión de las expectativas trabaja en todas direcciones. ¿Qué padre no ha descubierto una nueva fuerza de carácter, nuevos triunfos sobre la pereza, el miedo o la aprensión, al darse cuenta de que su hijo le estaba mirando? Como nos «ponemos a la altura de la situación», es bueno tener una vida llena de ocasiones y oportunidades para sacar lo mejor de nosotros mismos, ante los otros y *ante nosotros mismos*, y por lo tanto para aumentar las probabilidades de que este mejor yo tenga aún mejor aspecto en el futuro. (Ainslie, 2001, trata de forma particularmente sugerente esta dinámica.) La «presentación del yo en la vida cotidiana» (Goffman, 1959) es un baile interactivo de refinada coreografía (aunque sea en gran medida inconsciente), en el que no sólo tratamos de parecer mejores de lo que somos, sino que en el proceso sacamos lo mejor de los demás. No deberíamos jugar alegremente con los mecanismos que regulan estas prácticas, fruto de miles de años de evolución genética y cultural. Podría echar por el suelo una valiosa I+D. (*¡Detengan a ese cuervo!)* Por otro lado, si se hace con la debida comprensión y discernimiento, sí se pueden introducir algunas correcciones para reforzar o potenciar estos diseños, para evitar algunas oportunidades perdidas o algunas percepciones imprecisas. Es más, cierta intervención deliberada podría contribuir a eliminar aquellas variantes desafortunadas

de nuestras prácticas que pueden parecer contraproducentes. Aquí es donde entra en juego la capacidad que hemos desarrollado evolutivamente para la reflexión. Consideremos la sutil pero devastadora inclinación que descubrió la escritora afroamericana Debra Dickerson en su padre:

> Más tarde comprendí que él esperaba y necesitaba que los negros fracasaran, pues de otro modo no habría prueba alguna de la perfidia y la iniquidad de los blancos. Nunca comprendió que su fatalismo era una profecía autorrealizadora y autocontradictoria. Nunca consideró que debiera creer que los blancos eran superiores, a ningún nivel, porque consideraba que los negros no tenían ninguna oportunidad en la vida (pero probablemente lo habría atribuido al poder trascendente de la maldad innata de los blancos). Entre nosotros decimos que «el hielo de los blancos es más frío» para referirnos a los muchos de entre nosotros que no creen o valoran nada a menos que venga de los blancos. Cuanto peor es la situación de algunos negros, más mágicos les parecen los blancos, aunque sea una magia perversa.
>
> De este modo mi padre, igual que muchos otros negros, hacía el trabajo del opresor por él; y me enseñó a mí a hacer lo mismo. Fue en este momento cuando comencé a encerrarme en mí misma. Tal vez los blancos hubieran estado bien dispuestos a asumir ellos mismos la tarea, pero raramente tenían que hacerlo. Los blancos no tenían que poner barreras en mi camino, lo hacía yo misma al «aceptar» el lugar que tenía asignado al final de cualquier cola. El racismo y la desigualdad sistemáticas son fuerzas muy reales en nuestras vidas, pero también lo son el fatalismo y cierta exaltación perversa de la opresión (Dickerson, 2000, pág. 40).

¿Cuáles son las estructuras sociales a gran escala que mejor promueven la libertad y la distribuyen de manera más equitativa por el globo? ¿Qué combinación de normas explícitas y trucos sutiles tiene más probabilidades de modelar el entorno de un modo que promueva el crecimiento de las personas? En el capítulo 7 consideramos la idea de Robert Frank de que los problemas de compromiso y autocontrol se resuelven mutuamente en parte al favorecer la evolución de emociones como el enfado y el amor. Alian Gibbard desarrolla esta idea al especular sobre cómo podría afinar un «ingeniero psíquico» las disposiciones de la gente hacia el enfado, la culpa y otras emociones. El enfado, observa Gibbard, «es poderoso e inevitable, y a menudo ayuda a regular la acción en sentidos deseables» (Gibbard, 1990, pág. 298). Si bien «no podemos evitar enfadarnos, sean cuales sean nuestras normas» (pág. 299), algunas culturas no parecen dar ningún papel a la culpa. Esto plantea la cuestión de si no estaríamos mejor

sin esta emoción. Algunos deterministas radicales sostienen que no sólo no deberíamos lamentar la desaparición de la «genuina» libertad; también deberíamos celebrarlo como una suerte, puesto que sin la presunción de libertad podemos abandonar también las presunciones de responsabilidad moral, culpa y retribución, y vivir felices para siempre. He hecho todos los esfuerzos posibles por cortar la conexión que algunos imaginan que existe entre el determinismo y la responsabilidad, pero todavía podríamos preguntarnos, como Gibbard, si la moral es en sí misma algo que queremos preservar en nuestras sociedades. «En parte se trata de una cuestión pragmática: ¿acaso nos iría mejor sin esos sentimientos particulares, o sin normas que los gobernaran?» (pág. 295). La culpa y el enfado combinan bien: la culpa aplaca el enfado, y la amenaza de la culpa nos refrena de realizar actos que provocarían enfado. ¿Cómo tenderían a comportarse entre sí las personas en una sociedad que minimizara el papel tanto del enfado como de la culpa, o —con un heroico esfuerzo de ingeniería social— los anulara por completo? ¿Tal vez sería sabio por nuestra parte, por una razón u otra, reajustar la culpa y el enfado para que dejaran de estar en equilibrio, y uno dominara un poco sobre el otro? Los deterministas radicales dicen que el mundo sería un lugar mejor si consiguiéramos de algún modo deshacernos del sentimiento de culpa cuando causamos daño, y del enfado cuando nos lo hacen a nosotros. Pero no está claro que cualquier «cura» en este sentido no fuera peor que la «enfermedad». El enfado y la culpa tienen su razón de ser, y está profundamente arraigada en nuestra psicología.

Tal vez sea mejor, según Gibbard, promover unas condiciones que moderen la imperatividad de las normas que gobiernan estas emociones. Gibbard distingue entre un diseño «imperativo» y «moderado» de las normas morales. Las normas imperativas son muy exigentes, y provocan por lo tanto reservas privadas, hipocresía y desconfianza en los demás. Exigen un gran esfuerzo a la naturaleza humana y tienden a implicar «intimidaciones en cierto modo ineficientes». En su opinión, esto es un error de diseño, lisa y llanamente, del mismo modo que lo sería hacer demasiado sensible el volante de un coche, lo que haría que los conductores se pasaran al girar, luego al corregir el giro, luego al corregir la corrección, etc. Es un diseño inseguro y exige un esfuerzo innecesario al mecanismo sin conseguir los efectos deseados (pág. 306). Las normas moderadas, en cambio, son relativamente poco exigentes, un compromiso entre la prudencia y el interés personal que resulta más fácil de aceptar y, por lo tanto, más fácil de cumplir por parte de los individuos. Gibbard sugiere que el diseñador racional debería afinar las normas que gobiernan el enfado y la

culpa para que fueran más moderadas, un ajuste cultural que sacaría mayor partido a la naturaleza, sin enfrentarse a ella.

Consideremos el caso del «pensador privado», un individuo a quien Gibbard presenta enfrentado a la competición entre sus fines egoístas y la llamada general a la benevolencia, o la moral. Supongamos que en el debate público el pensador se ve arrastrado a expresar su asentimiento con relación a ciertas normas públicas, a pesar de lo cual mantiene sus reservas con respecto a ellas y se pregunta si debe hacer realmente lo que le piden cuando podría no hacerlo y salir igualmente bien parado. Tal vez esté familiarizado con la tesis de Robert Frank de que, desde un punto de vista prudencial, sale a cuenta *ser* bueno para *parecer* bueno, pero tal vez juegue con la idea de que él pueda ser la excepción a la norma. Ha aceptado ayuda de sus amigos pero es capaz de preguntarse hasta qué punto le sale a cuenta este intercambio, si exige reciprocidad por su parte. ¿No se habrá visto empujado a convertirse en un buen ciudadano meramente por las exigencias circunstanciales de la conversación? La resolución de este conflicto depende en gran medida de la atmósfera social:

> **Si la integridad moral es la mejor forma de promover sus fines egoístas, la ambivalencia está resuelta. Es improbable que se dé esta circunstancia con una moral imperiosa; con una moral moderada, parece más plausible. [...] Lo que caracteriza a una moralidad moderada es que se alia con los suficientes motivos ajenos a ella como para prevalecer en la mayoría de los casos (para prevalecer entre las personas actuales, con todo lo que las une y las separa, con sus motivaciones normativas y sus apetitos, sentimientos, impulsos y aspiraciones) (Gibbard, 1990, pág. 309).**

Los ingenieros, igual que los políticos, se dedican al arte de lo posible, lo cual exige por encima de todo realismo respecto a cómo son actualmente las personas, y cómo llegaron a ser de esta manera. Las teorías éticas que se niegan a doblegarse a los hechos empíricos sobre la humanidad están condenadas a generar fantasías que tal vez puedan tener cierto interés estético, pero que no deberían tomarse en serio como recomendaciones prácticas. Como todo lo que ha creado la evolución, somos una caja de trucos inventados de manera más o menos oportunista, y nuestra moral debería estar basada en esta idea. Los filósofos han tratado a menudo de establecer una moral hiperpura, ultrarracional, no contaminada por la «compasión» (Kant) ni el «instinto», ni por disposiciones, pasiones o emociones animales. Gibbard lanza una mirada pragmática sobre los materiales que tenemos para trabajar y propone hacer, como ingenie-

ros, lo que la Madre Naturaleza ha hecho siempre: trabajar a partir de lo que hay.

AUTONOMÍA, LAVADO DE CEREBRO Y EDUCACIÓN

> ***Tomarse** a uno mismo como agente racional es asumir que la propia razón tiene una aplicación práctica o, lo que es lo mismo, que se tiene voluntad. Es más, uno no puede asumir esto último sin presuponer de entrada la idea de la libertad, razón por la cual sólo se puede actuar, o pensar que se actúa, bajo esta idea. Constituye, como si dijéramos, la forma del pensamiento de uno mismo como agente racional.*
>
> HENRY **A.** ALLISON, **«We Can Act Only under the Idea of Freedom»**

La explicación que he esbozado del arte de hacerse a uno mismo lo hace depender de un inquietante número de manipulaciones inconscientes o subliminales, además del ejercicio de la «razón pura». ¿No socava este proceso mismo el concepto de un yo responsable? Esta cuestión ha sido ampliamente explorada por Alfred Mele en *Autonomous Agents* (1995). Mele sostiene que más allá del mero autocontrol hay lo que se llama la *autonomía,* que distingue a su vez de la *heteronomía,* la cual se produce cuando un agente capaz de ejercer autocontrol se encuentra sin embargo bajo el control (parcial) de otros. Mele propone un Principio de Responsabilidad por Defecto: si ninguna otra persona es responsable de que usted se encuentre en el estado A, lo es usted. Esta maniobra corta limpiamente el regreso al infinito que tanto temía Kane; nos permite pasarles la pelota a los «lavadores de cerebros» (si es que se ha cruzado con alguno en el pasado), pero no a la «sociedad» en general o a un entorno sin agentes. Sólo si le han estado manipulando para sus propios fines unos agentes dotados de voluntad y previsión está usted absuelto de responsabilidad personal por las acciones emprendidas por su cuerpo; no son acciones suyas, sino de aquellos que le han lavado el cerebro. Hasta aquí ningún problema, pero cabe decir que también los educadores diseñan sus interacciones con nosotros con objeto de perseguir sus propios fines, en particular el fin de convertirnos en agentes morales dignos de confianza. ¿Cómo podemos distinguir entre la buena educación, la propaganda dudosa y el pérfido lavado de cerebro? ¿Cuándo nos estamos benefician-

do de la ayuda de nuestros amigos, y cuándo se están aprovechando de nosotros?

El término que usa Mele para referirse al lavado de cerebros es el de «ingeniería de valores» y habla en términos muy negativos de dicha ingeniería cuando «elude» la capacidad de las personas para controlar su vida mental (Mele, 1995, págs. 166-167). Tal como hemos visto en capítulos anteriores, el autocontrol que tenemos sobre nuestra vida mental es en todo caso limitado y problemático, de modo que no es ninguna sorpresa que tengamos problemas para distinguir la ingeniería que elude nuestras capacidades de la ingeniería que las explota de una manera tolerable o deseable. Para poner de relieve la diferencia entre autonomía y heteronomía, Mele propone algunos experimentos mentales sobre dos agentes separados por diferencias mínimas, Ann y Beth. Supongamos, para empezar, que Ann es genuinamente autónoma (sea lo que sea lo que eso signifique). Suerte para Ann. Luego supongamos que Beth es igual que Ann, su gemela idéntica desde el punto de vista psicológico, podría decirse, pero sin que ella lo sepa le han practicado un lavado de cerebro para dejarla en su estado psicológico actual, tal vez sólo aparentemente envidiable. Beth tiene las mismas disposiciones que Ann; es igual de abierta y poco obsesiva que ella, tan flexible aunque decidida como Ann, pero su aparente autonomía, según Mele, es falsa. Viene a ser como la perfecta falsificación de un dólar, fácilmente intercambiable por una Coca-cola y algo de cambio, y sin embargo inautèntico en un sentido importante, un sentido moral.

Los experimentos mentales que proponen unas estipulaciones tan extremas —y poco realistas— tienden a confundir la imaginación del filósofo, por lo que es importante tocar todos los botones, variar todas las estipulaciones en un sentido y en otro, para ver de dónde proceden realmente las intuiciones. Normalmente, en el mundo real, la importancia que pueden tener las cuestiones históricas (en este caso, la educación de Ann frente al lavado de cerebro de Beth) es que supongan alguna diferencia en la disposición o el carácter que a su vez dará lugar a diferencias en el comportamiento futuro. Esto es precisamente lo que hemos descartado en el caso imaginario, pero ¿podemos admitir esta estipulación tal como está formulada? Los experimentos mentales relativos a lavados de cerebro son endémicos en las discusiones filosóficas acerca de la libertad, y un elemento rutinario —aunque raramente comentado— en los mismos es la estipulación de que la víctima no recuerda nada de la intervención. Veamos lo que ocurre si cambiamos esto. Supongamos, como Mele (1995, pág. 169), que posteriormente se informa a Beth de su historia secreta y se

le ofrece la posibilidad de pedir que le deshagan su lavado de cerebro. Si ella da su conformidad retrospectivamente, ¿qué *valor* tiene este acto?, ¿es *a partir de ahora* un agente autónomo? Tal vez nuestras intuiciones vacilen en este punto, puesto que el estado de Beth cuando «da su conformidad» es también (por hipótesis) un producto de su lavado de cerebro anterior. Tal vez usted quisiera objetar que ella ha sido *diseñada para dar su conformidad a su propio diseño*, lo cual se convierte en un gesto vacío por su parte. Pero no está tan claro. Consideremos la diferencia que podría introducir el tiempo. Supongamos que esperamos unos años antes de informarle de su historia secreta y le damos gran cantidad de experiencia en el ancho y diverso mundo de las decisiones morales. Como Beth es tan abierta y cognitivamente flexible como Ann (por hipótesis), esta experiencia será tan intensa y valiosa para ella como lo sería para Ann y, por lo tanto, sería tan válida para *fundar* un consentimiento en su caso como en el de Ann. Podemos llevar aún más lejos esta línea de razonamiento si ahora suponemos que introducimos la misma variación en el caso de Ann: le decimos a *ella* que ha sido víctima de un lavado de cerebro (lo que es mentira). Ella reflexiona sobre este dato y decide que aprueba su manera de ser: ¡qué otra cosa queremos que haga! Después de todo, es *efectivamente* autónoma (sea lo que sea lo que eso signifique). ¿Vale algo más su acto que el de Beth? No veo ningún motivo para ello. Yendo más al fondo de la cuestión, tal vez sienta usted alguna inclinación a suponer que al mentirle a Ann le hemos causado algún perjuicio en su departamento de autonomía (suponiendo que se crea nuestra mentira, claro). ¿Por qué? Porque ahora está gravemente desinformada respecto a su pasado, haga o no uso de esta información distorsionada en sus decisiones. (Y resulta fácil imaginar que esta desinformación podría alterar todas sus reflexiones posteriores sobre temas morales.)

Pero recordemos que Beth estaba también radicalmente desinformada antes de que le contáramos lo de su lavado de cerebro. ¿No es así? Mele no entra en esta cuestión, pero seguramente se le ocultó que le habían practicado un lavado de cerebro; presumiblemente parte de su parecido psicológico con Ann antes de que se le revele su secreto es un conjunto asombrosamente rico de falsas pseudomemorias de una educación moral propiciadora de autonomía que nunca tuvo lugar. ¿Cómo puede mantenerse si no la estipulación de que sea la gemela psicológica de Ann?

¿No podría ser, entonces, que las marcas definitorias del lavado de cerebro fueran simplemente la falsedad y la ocultación? En la medida en que digamos la verdad (lo que se considera la verdad en el momento de

decirlo) y evitemos engañar a las personas, mientras las dejemos en un estado desde el que puedan realizar una evaluación independiente de su situación al menos tan buena como la que pudieran hacer antes de nuestra intervención, no les estamos practicando ningún lavado de cerebro, sino que las estamos educando. La idea de que la propia historia pueda suponer una diferencia moralmente importante sin que introduzca ninguna diferencia para la propia competencia futura no queda apoyada en absoluto por los experimentos mentales de Mele. Su comparación con la falsificación perfecta de un dólar es instructiva en este sentido. Las falsificaciones importan por sus efectos sobre las creencias y deseos del pueblo respecto a la fiabilidad de su dinero, pero esos efectos son de carácter general, no afectan para nada a los billetes concretos. La identificación y retirada de falsificaciones perfectas de la moneda en circulación sería un proyecto inútil, puesto que la diferencia entre un dólar auténtico y una falsificación perfecta es *(ex hypothesi) un* hecho histórico inerte. La creencia de que *hay* una gran cantidad de falsificaciones perfectas en la moneda de curso legal podría trastornar la economía al debilitar la confianza en el control del gobierno sobre su política monetaria, pero no tendría ningún sentido reunir los billetes falsificados y destruirlos (en lugar de reunir y destruir cualquier conjunto de dólares en circulación).

Volvamos al caso de Ann y Beth. Si Beth llega a conocer la verdad sobre su lavado de cerebro, ello lanzará sin duda sombras inquietantes por toda su psique, y quién sabe qué efectos tendrá sobre su competencia moral. Pero Ann experimentará exactamente lo mismo si se le explica convincentemente la misma «verdad» sobre ella misma. El perjuicio es igual en ambos casos. Y si la autonomía de Ann depende de la verdad de sus propias creencias acerca de su pasado, entonces el problema de Beth es simplemente que le han mentido, no que le hayan llevado hasta un envidiable estado disposicional a través de técnicas de «ingeniería de valores». Nótese, por cierto, lo que esto presagia para cualquier doctrina que pretenda defender el *¡Detengan a ese cuervo!* sobre la base de que la gente está mejor si no sabe la verdad: «Tuvimos que destruir la autonomía de la humanidad para salvarla». No es un eslogan muy atractivo.

El agente genuinamente moral es racional, capaz de ejercer autocontrol, y no es víctima de ninguna desinformación grave. La repugnancia intuitiva que sentimos ante las «píldoras morales» y los «lavados de cerebro», frente a la vieja y conocida educación moral, se debe tal vez a una vaga apreciación de la completa imposibilidad de que haya algún tratamiento abreviado de este tipo que pueda preservar realmente nuestra in-

formación, flexibilidad y apertura, que según nuestra experiencia depende de una buena educación. No veo por qué tomar *conscientemente* una píldora para mejorar el propio autocontrol ha de ser algo más subversivo para la propia autonomía que fomentar deliberadamente un cierto autoengaño respecto a las propias capacidades. El hecho de que usted esté dispuesto, como adulto responsable, a manipularse a sí mismo de este modo y a asumir los efectos tanto prospectivos como retrospectivos de dicha manipulación puede servir como un test relativamente bueno para saber si está justificado que manipule del mismo modo a sus hijos. En Lake Wobegon, la ciudad mítica de Garrison Keillor, «todos los niños están por encima de la media», y este mito feliz les hace mejores de lo que serían de otro modo (mientras no lleguen a engañarse demasiado a este respecto). Ciertamente es mejor que creer que el hielo del hombre blanco es más frío.

Los filósofos han explorado también otro punto de vista sobre la autonomía, siguiendo los pasos del influyente artículo de Harry Frankfurt: «Freedom of the Will and the Concept of a Person» (1971). Frankfurt propuso la idea de que una persona —un agente adulto y responsable— se distingue de un animal o de un niño por tener una psicología más compleja: en particular, por tener deseos de orden superior. Puede darse el caso de que una persona quiera una cosa y, sin embargo, *quiera* querer otra cosa, y actúe de acuerdo con este deseo de segundo orden. Tal capacidad para reflexionar sobre los deseos que uno descubre en uno mismo, para luego asumirlos o rechazarlos, no es sólo un signo de madurez, según Frankfurt; es el criterio que define la personalidad. Esta idea intuitivamente atractiva ha demostrado bastante resistencia a entrar en una formulación que no caiga en regresiones o contradicciones, y un interesante intento relativamente reciente de David Velleman en este sentido subraya tanto el papel que debe reservarse al razonamiento como el requisito de no hacernos demasiado pequeños: «La función del agente, según Frankfurt, es reflexionar sobre los motivos que compiten por gobernar su comportamiento, y determinar el resultado de la competición al optar por algunos de sus motivos en lugar de otros» (Velleman, 1992, pág. 476). ¿Cómo puede una persona optar a favor o en contra de algunos de sus propios motivos?

Consideremos las siguientes diferencias entre dos monjes católicos: uno abraza ardientemente su voto de celibato y triunfa sobre su constitución genética gracias a la fuerza de su voluntad; el otro es igualmente célibe, pero ve su catolicismo como una adicción. Considera que le han lavado el cerebro, que es la víctima de unos memes invasores, pero no consigue convencerse para dar el salto y abandonar los principios que le han

enseñado. Ciertamente hay personas reales que entran en estas dos categorías en muchos campos, pero ¿en qué consiste primariamente la diferencia? Ambos monjes están fuertemente *motivados* por los principios del catolicismo, pero uno se identifica sinceramente con su religión, mientras que el otro no. La identificación no depende de que un ego cartesiano o un alma inmaterial acepte unos memes y rechace otros; la entidad que asume o no dichos memes tiene que ser también algún tipo de meme o estructura cerebral compleja. Pero ¿cómo podemos aceptar una estructura de este tipo, una especie de agente dentro del agente capaz de optar por un bando u otro, sin caer otra vez en los misterios cartesianos sobre una *res cogitans* independiente que ejerce el papel de jefe, o por lo menos el de juez o guardia de tráfico, dentro de la caótica competición que tiene lugar en el cerebro? Velleman ofrece un ejemplo en este sentido que recuerda algunos de los experimentos de Daniel Wegner, en los que la acción viene gobernada por una conspiración sumergida, parcial o incluso enteramente inconsciente de motivos, razones, reconocimientos y demás:

> **Supongamos que acudo a una cita largo tiempo esperada con un viejo amigo para resolver una diferencia menor, pero que en el curso de nuestra conversación me siento provocado por algunos comentarios casuales suyos, levanto la voz y soy cada vez más cortante hasta que al final nos separamos en términos nada amistosos. Una reflexión ulterior me lleva a comprender que las ofensas acumuladas habían cristalizado en mi mente, durante las semanas previas al encuentro, hasta convertirse en una resolución de terminar con nuestra amistad por aquella cuestión, y que fue esta resolución lo que dio el tono cortante a mis comentarios [...]. Pero ¿debo pensar necesariamente que tomé la decisión, o que sólo la ejecuté? [...] Cuando mis deseos y creencias engendran la intención de cortar una amistad y cuando esta intención hace que adopte un tono agresivo, están ejerciendo los mismos poderes causales que poseen en circunstancias corrientes, y, sin embargo, lo hacen sin ninguna contribución por mi parte (Velleman, 1992, págs. 464-465).**

¿Qué habría cambiado si hubiera existido dicha contribución? Tal como observa Velleman, un agente debe ser más que un punto matemático, puesto que

> **cuando opta por alguno de sus motivos, le confiere una fuerza adicional a la que tienen por sí mismos —y por lo tanto distinta de ella—. [...] ¿Qué evento o estado mental podría desempeñar esta función de dirigir siempre pero no someterse nunca a este escrutinio? Sólo puede ser otro motivo que dirija el propio pensamiento práctico (págs. 476-477).**

Y este motivo sólo puede ser, tal como dijo Kant hace largo tiempo, el respeto mismo por la razón: «Lo que anima el pensamiento práctico es una preocupación por actuar de acuerdo con razones» (pág. 478). ¿Y de dónde procede esto? De la crianza, que hace participar al hijo en la práctica de pedir y dar razones. La función de la conciencia aquí es precisamente llevar los asuntos al terreno de la deliberación y la reflexión, donde *con el tiempo* se puedan considerar y negociar las razones en pro y en contra. Pero, ¿qué pasa con aquellos jesuítas que (según se dice) decían tener bastante con los primeros siete años de un niño para hacer que se identificara con la fe? ¿Es eso adoctrinamiento o educación? Pienso que el planteamiento que estoy esbozando sale antes reforzado que debilitado por dar la posibilidad de que ambos monjes católicos puedan tener razón; el primero no tiene por qué engañarse a sí mismo en su creencia de que tiene la autonomía necesaria para asumir su decisión y cumplirla, y el segundo puede estar justificado al lamentarse por su adoctrinamiento, y, sin embargo, las diferencias entre sus respectivas educaciones pueden ser mínimas. Las personas son seres extraordinariamente complicados, y lo que es bueno para una puede ser muy perjudicial para otra. (Lo mismo puede decirse del Ritalin, por supuesto; muchos niños a quienes se les ha prescrito no deberían estar tomándolo.) ¿Cuál es entonces el importante papel que desempeña este yo? El yo es un sistema al que se *atribuye* la responsabilidad, a lo largo del tiempo, con objeto de poder confiar en que esté allí para *asumir* la responsabilidad, para que siempre haya alguien que responda cuando se plantean cuestiones de responsabilidad. Kane y los otros tienen razón al buscar un lugar en el que termine la cadena. Simplemente han estado buscando una cosa equivocada.

Capítulo 9

La cultura humana ha hecho posible la evolución de mentes lo bastante poderosas como para captar las razones que hay detrás de las cosas y hacerlas suyas. No somos agentes perfectamente racionales, pero el mundo social en el que vivimos genera procesos de interacción dinámica que requieren y al mismo tiempo permiten la renovación y la suscripción de nuestras razones, y nos convierten en agentes capaces de asumir la responsabilidad por sus acciones. Nuestra autonomía no depende de nada parecido a una suspensión milagrosa de los procesos causales, sino más bien de la integridad de los procesos educativos y de intercambio de conocimiento.

Capítulo 10

Las auténticas amenazas a la libertad no son metafísicas, sino políticas y sociales. A medida que vayamos conociendo las condiciones que hacen posible la toma de decisiones en los seres humanos, deberemos diseñar y acordar sistemas jurídicos y de gobierno que no sean rehenes de falsos mitos sobre la naturaleza humana, que sean sólidos frente a posibles descubrimientos científicos y avances tecnológicos. ¿Somos más libres de lo que queremos ser? Ahora tenemos más poder que nunca para crear las condiciones bajo las cuales viviremos nosotros y nuestros descendientes.

NOTAS SOBRE FUENTES Y LECTURAS COMPLEMENTARIAS

Don Ross me ha hecho ver que el análisis de Skyrms no posee la suficiente generalidad y que para encontrarla hay que ir al reciente (y terriblemente matemático) *Game Theory and the Social Contract*, vol 2, *Just Playing* (1998), de Ken Binmore.

En *Elbow Room* (Dennett, 1984) puede encontrarse una versión anterior de mi explicación gradualista sobre cómo nos aupamos a la moral, en el capítulo 4, «Self-made Selves». La exposición actual complementa y no rescinde en ningún sentido la anterior.

El artículo de Peter Súber «The Paradox of Liberation» (no publicado, pero disponible en Internet en http://www.earlham.edu/~peters/writing/liber.htm), de 1992, ha sido una gran fuente de intuiciones para mí, y me ha proporcionado también las magníficas citas de James Branch Cabell y de Alcohólicos Anónimos que aparecen como epígrafes.

Véase *The Nurture Assumption* (1998), de Judith Harris, para encontrar un estudio basado en un amplio espectro de variables psicológicas según el cual los niños reciben una influencia más fuerte de sus compañeros que de sus padres.

The Moral Animal (1994), de Robert Wright, contiene algunos comentarios sobre *La presentación de la persona en la vida cotidiana,* de Goffman, en el capítulo dedicado al engaño y al autoengaño.

Véase mi artículo «Producing Future by Telling Stories» (1996c), dedicado al papel que desempeñan los cuentos de hadas en la construcción de agentes dignos de confianza. La obra de Victoria McGeer ha sido la fuente principal de mis comentarios sobre el papel del entorno en este sentido. También es relevante la abundante literatura relativa a la «teoría

de la mente infantil», a la que puede darse un repaso a través de Astington, Harris y Olson, 1988; Baron-Cohen, 1995; y Baron-Cohen, Tager-Flusberg y Cohen, 2000.

Aquellos que quieran investigar los atractivos y los problemas del determinismo radical y posturas similares pueden consultar «Ethics without Free Will» (1990), de Michael Slote; *La máquina de los memes* (1999), de Susan Blackmore; y *Living without Free Will* (2001), de Derk Pereboom.

Para leer más sobre experimentos mentales extremos que nos reclaman que nos tomemos en serio fantasías tales como las pildoras morales y los lavados de cerebro que no dejan marca alguna, véase mi artículo «Cow-sharks, Magnets, and Swampman» (Dennett, 1996b).

Sobre Hume, véase «Natural and Artificial Virtues: A Vindication of Hume's Scheme» (1996), de David Wiggins.

Capítulo 10

El futuro de la libertad humana

¿Dónde nos llevará todo eso? No hay motivo de inquietud más poderoso en relación con la libertad que la imagen de las ciencias físicas sumergiendo todas nuestras acciones, buenas o malas, en el corrosivo caldo de la explicación causal, desguazando el alma pieza por pieza hasta que no quede nada que elogiar o condenar, que honorar, respetar o amar. O eso piensa mucha gente. Y por lo tanto se esfuerzan por levantar una barrera tras otra, algún tipo de doctrina absolutista diseñada para mantener a raya esas deletéreas ideas. Pero es una estrategia condenada al fracaso, una reliquia del milenio pasado. El avance de nuestro conocimiento de la naturaleza nos ha enseñado que tales bastiones no hacen sino posponer la catástrofe, y a menudo la hacen todavía peor. Si quiere usted vivir en la costa, será mejor que esté preparado para mudarse cuando la playa cambie de lugar, tal como efectivamente hacen las playas, a un ritmo lento pero seguro. Los espigones pueden «salvar» la costa sólo al precio de destruir algunos de los rasgos que hacían de la costa un lugar tan agradable para vivir. Lo más inteligente es estudiar la situación y ponerse de acuerdo en unas cuantas normas básicas que determinen a qué distancia de la costa puede construirse una casa. Pero los tiempos cambian, y las políticas que funcionaron durante décadas o siglos pueden convertirse en obsoletas y necesitar una revisión. A menudo se dice que debemos trabajar con la naturaleza, no contra ella, pero eso no es más que la retórica de la moderación; cada artificio humano obstaculiza o redirige alguna tendencia natural; el truco es llegar a saber lo bastante sobre la manera en que se articulan las pautas naturales para que nuestra interferencia en ellas produzca los resultados deseados.

MANTENERSE FIRMES ANTE LA PROGRESIVA EXCULPACIÓN

A medida que vayamos aprendiendo más acerca de la constitución de la mente de las personas, las presunciones implícitas en nuestras instituciones de mérito y culpa, castigo y tratamiento, educación y medicación deberán ajustarse a los hechos tal como los vayamos conociendo, pues una cosa está clara: las instituciones y las prácticas basadas en falsedades evidentes son demasiado endebles como para merecer nuestra confianza. Pocas personas estarán dispuestas a apostar su futuro a un mito frágil cuyas grietas pueden ver ellos mismos. En realidad, nuestras actitudes en estas materias han ido cambiando gradualmente a lo largo de los siglos. Actualmente no dudamos en exculpar o mitigar la pena en muchos casos que nuestros antepasados habrían tratado con mucha más dureza. ¿Es eso una señal de progreso o de que nos estamos volviendo todos blandos con el pecado? Para los miedosos, esta revisión parece un síntoma de erosión, y para los optimistas tiene todo el aspecto de un progreso, pero también existe una perspectiva neutral desde la que contemplar el proceso. A ojos de un evolucionista se presenta como un frágil equilibrio, necesariamente transitorio, que constituye el resultado relativamente estable de una serie de innovaciones y contrainnovaciones, ajustes y contraajustes, en una carrera armamentista que genera al menos un tipo de progreso: un avance del autoconocimiento, un aumento de la capacidad de análisis de quién somos y lo que somos, y de lo que podemos y no podemos hacer. Y desde este autoconocimiento definimos y redefinimos nuestras conclusiones sobre lo que debemos hacer.

Volvamos a una pregunta del capítulo 9 que quedó sin respuesta: ¿qué condiciones debe cumplir una persona para que podamos considerarla genuinamente culpable de sus fechorías? y ¿hay alguien que las cumpla realmente? Nadie es perfecto, y por otro lado la idea de un perfecto *malhechor* está al borde de caer en la contradicción, tal como ya Sócrates apreció. ¿No tiene que haber *algún tipo* de desajuste en una persona que se disponga a hacer el mal a sabiendas? ¿Dónde debemos marcar la línea entre las diversas patologías exculpatorias —él no lo sabía, él no podía controlarse— y la gente que hace el mal «por su propia y libre voluntad», a sabiendas de lo que hace? Si ponemos el umbral demasiado alto, todo el mundo queda impune; si lo ponemos demasiado bajo, terminamos por castigar a cabezas de turco. Las diversas propuestas libertaristas en este sentido se quedan muy lejos de su objetivo: causación por el agente en términos francamente misteriosos, indeterminación cuántica en la facul-

tad de la razón práctica, levitación moral a cargo de almas inmateriales y otros marionetistas espectrales; en el mejor de los casos estas doctrinas pueden servir para que apartemos nuestra atención de un problema difícil y la concentremos, en cambio, en un misterio convenientemente insoluble. De modo que volvamos al problema: *¿dónde* debemos marcar la línea, y qué impide que ésta se retire indefinidamente ante la presión de la ciencia?

Imaginemos que tratamos de diseñar una prueba de aptitud para medir la flexibilidad mental, la cultura general, la comprensión social y el control de los impulsos, que son lo que podríamos considerar razonablemente los requisitos mínimos para la agencia moral. Dicha prueba podría hacer operativo el ideal implícito en nuestra concepción de la responsabilidad: los adultos normales cumplen con los requisitos, y son algo que se cumple o no se cumple. Podríamos diseñarla de modo que tuviera un «efecto techo»: no se pueden conseguir más de 100 puntos de 100, y la mayoría de las personas obtienen 100. (No tenemos ningún interés en las diferencias de competencia que puedan existir por encima del umbral. El obtuso Smith tal vez no sabía lo que hacía con la misma claridad que su cómplice, el brillante Jones, pero lo sabía lo bastante como para ser responsable de ello.) La razón de una política como ésta es clara y conocida, y parece funcionar bien en aplicaciones sencillas como el permiso de conducir. Es preciso tener 16 años (o 15, o 17...) y pasar una prueba de aptitud y conocimiento de las reglas. A partir de entonces se concede a la persona en cuestión la libertad de circular por la carretera y se la trata como cualquier otro conductor. Esta política está abierta a revisiones a medida que vayamos aprendiendo más cosas acerca de sus efectos sobre la seguridad vial: restricciones nocturnas, períodos de aprendizaje, excepciones motivadas por discapacidades reconocibles u otras circunstancias especiales que puedan considerarse desde un equilibrio coste-beneficio entre la maximización de la seguridad y la maximización de la libertad.

Podemos reconocer un proceso similar oculto tras los debates sobre los motivos de exculpación o mitigación de la responsabilidad en general. A medida que vamos aprendiendo más cosas sobre las posibles causas de discapacidad parcial y sus efectos, descubrimos razones para reubicar a ciertos individuos en relación con el umbral antes mencionado, habitualmente —aunque no siempre— en la dirección de exculpar a algún colectivo de personas cuya culpabilidad no ofrecía dudas hasta el momento. Esto puede generar la impresión de que nos encontramos ante un umbral en constante retirada, pero es preciso que examinemos la cuestión de manera

desapasionada. Podemos introducir perfectamente importantes revisiones en nuestras políticas relativas a quién encarcelamos y a quién ponemos bajo tratamiento, por ejemplo, sin que ello suponga ninguna revisión de nuestras premisas filosóficas de fondo. Después de todo, no cambiamos nuestros conceptos de culpa e inocencia cada vez que descubrimos que ha sido un error encarcelar a una persona determinada. Sacamos a esa infortunada persona del conjunto de aquellos que consideramos culpables, pero no cambiamos los demás criterios establecidos. Precisamente porque suscribimos nuestra noción estándar de culpa reconocemos que esta persona no era culpable en realidad. De modo parecido, podemos exceptuar a una *categoría* de individuos del conjunto de los que consideramos responsables, sin que ello suponga ningún cambio —en particular, ninguna «erosión»— de nuestro concepto de responsabilidad moral. Simplemente habríamos aprendido que hay menos personas moralmente responsables en nuestra sociedad de las que antes pensábamos.

Pero el temor subsiste: «¿Dónde se detendrá el proceso?». ¿No estaremos avanzando hacia una sociedad cien por cien «medicalizada», en la que nadie será responsable, y todo el mundo será víctima de algún lamentable defecto en su constitución (tenga su origen en la naturaleza o en la crianza)? No, porque hay fuerzas —nada misteriosas ni metafísicas, sino fuerzas políticas y sociales fácilmente comprobables— que se oponen a esta tendencia, y son análogas a las fuerzas que impiden que la edad para conducir se eleve, por ejemplo, a los 30 años. La gente *quiere* ser responsable. Los beneficios que ello supone para un ciudadano respetado en una sociedad libre son tantos y tan valiosos que siempre hay una fuerte presunción en favor de la inclusión. La culpa es el precio que pagamos por ganarnos el crédito de los demás, y lo pagamos gustosamente en la mayoría de los casos. Lo pagamos caro, y aceptamos el castigo y la humillación pública a cambio de la oportunidad de que nos vuelvan a admitir en el juego después de que nos hayan atrapado en alguna transgresión. Y, por lo tanto, la mejor estrategia para mantenerse firmes frente a la progresiva exculpación está clara: proteger y potenciar el valor de los juegos a los que uno puede jugar cuando es un ciudadano respetado. Es la erosión de estos beneficios, no el avance de las ciencias humanas y biológicas, lo que pondría en peligro el equilibrio social. (Recordemos el cínico eslogan que acompañó el declive y el colapso final de la Unión Soviética: ellos fingen que nos pagan y nosotros fingimos que trabajamos.)

Sin duda siempre tendremos fuertes tentaciones de hacernos muy pequeños, externalizar las causas de las propias acciones y rechazar la res-

ponsabilidad, por lo que la mejor respuesta será que nos hagan una oferta que nadie pueda rechazar: si usted quiere ser libre, debe *asumir* su responsabilidad. Pero ¿qué sucede con los pobres desgraciados que no saben hacerse cargo de sus vidas y cuya capacidad para resistir a la tentación es tan limitada que tienen prácticamente asegurada una vida de transgresiones y castigos? ¿Acaso no es injusto con ellos, acaso no es una oferta coercitiva que sólo en apariencia permite una elección libre? Estas personas son realmente incapaces de cumplir su parte del trato, y se les castiga por ello. Tal vez sean unos buenos cabezas de turco y el ejemplo que sentamos con ellos sirva para mantener viva la imagen del castigo que reprime a muchos otros agentes dotados de una capacidad de autocontrol levemente superior, pero ¿acaso no es esto una política manifiestamente injustificable? Al fin y al cabo, «no podían hacer otra cosa». Hay un sentido de esta gastada frase que resulta aplicable a este contexto, pero tal como veremos no es el que preocupa a los incompatibilistas.

La dinámica del proceso de negociar umbrales resulta tal vez más visible en los casos extremos que se presentan ocasionalmente ante el público. ¿Qué debemos hacer, por ejemplo, en el caso de los pederastas? La tasa de reincidencia es elevadísima —son realmente perros viejos incapaces de aprender nuevos trucos, según parece— y el daño que hacen si se les deja en libertad también es terrible (Quinsey y otros, 1998). Existe, sin embargo, un tratamiento que, según han demostrado los estudios, resulta efectivo para dotar a los pederastas del autocontrol necesario para que sea seguro devolverlos a la sociedad (bajo cierta supervisión ulterior): la castración. Un severo remedio para una severa enfermedad. ¿Es justificable? ¿Es un «castigo cruel e inaudito»? Es importante señalar que muchos pederastas condenados aceptan voluntariamente la castración, como alternativa altamente preferible a un encarcelamiento indefinido. (Se escuchan menos quejas respecto al cruel e inaudito castigo de dejar en libertad a los violadores en medio de una comunidad de ciudadanos debidamente aterrorizados e indignados dispuestos a formar piquetes para expulsar al peligroso individuo.) La cuestión está lejos de quedar resuelta y se ve complicada por muchos factores. La castración produce su efecto principal al detener el flujo de testosterona en el cuerpo, y esto es algo que se puede conseguir por medios químicos o quirúrgicos. La castración química requiere repetidas inyecciones y es en general reversible, pero las drogas tienen algunos efectos secundarios perjudiciales; la castración quirúrgica no es directamente reversible, pero su efecto principal sobre el comportamiento puede evitarse mediante la autoadministración de testosterona (si uno realmente

quiere hacerlo). Pero ¿por qué habría de querer alguien hacer una cosa así? (Véanse, por ejemplo, Prentky, 1997, y Rosler y Witztum, 1998.)

El efecto simbólico de la castración es sin duda parte del motivo de todo el revuelo suscitado por la cuestión. Si la extirpación quirúrgica del apéndice, por decir algo, tuviera efectos igualmente positivos sobre el autocontrol de aquellos que se sometieran al tratamiento, cuesta creer que la opción encontrara una oposición tan vehemente. Sé por experiencia que sacar esta cuestión en el presente contexto hará que algunos lectores se lleven las manos a la cabeza. «¡Termina defendiendo la castración!» No, he presentado dicha política como una alternativa seria, pero no he expresado ninguna opinión respecto a su conveniencia última. Después de todo, es posible que estemos cerca de encontrar otro tratamiento mejor y menos severo. Es más, supongamos, por mor del argumento, que la tasa de reincidencia de los pederastas fuera del 50 % (lo que no se aleja mucho de la realidad), y supongamos que muchos pederastas aceptaran voluntariamente la castración como el precio que están dispuestos a pagar por la libertad. Aproximadamente la mitad de estas castraciones serían «innecesarias»: los sujetos no hubieran vuelto a reincidir. El problema es que no podemos (todavía) reconocerlos por adelantado. Pero presumiblemente esta situación mejorará a medida que avance nuestro conocimiento. ¿Qué deberíamos hacer mientras tanto? Hay razones convincentes en contra de la castración, y razones convincentes a favor de ella. Uso la castración sólo como ejemplo, e invito a los lectores a reflexionar sobre lo fuerte que pueda ser en ellos el impulso de cerrar su mente y fiarse sólo de su «corazón» ante una propuesta tan «incalificable» como ésta. Esto es parte del problema. Algunas personas están tan convencidas de que tratan de llevarlas por el camino de la perdición que no se permiten pensar sobre estas cuestiones. Se supone que los filósofos están por encima de esta clase de presiones, que son espectadores desapasionados de toda opción concebible, aislados en sus torres de marfil, pero eso no es más que un mito. En realidad, a los filósofos les gusta mucho más hacer el papel de vigías, y lanzar la voz de alarma ante cualquier catástrofe vagamente imaginada antes de que ésta llegue a tener la menor oportunidad de presentarse.

La castración es un ejemplo útil porque pone al descubierto las inconsistencias en los planteamientos de ambos bandos. Hay quien busca ávidamente una medicación que le ayude a seguir la dieta o que le controle la presión sanguínea que no es capaz de controlar mediante un ejercicio adecuado, al tiempo que niega a otros el derecho a usar los mismos refuerzos

tecnológicos para afianzar o complementar la voluntad frente a otras tentaciones. Si en su caso es racional y responsable reconocer sus propias debilidades y adoptar las medidas disponibles para reforzar su propio autocontrol, ¿cómo puede criticar las mismas medidas en el caso de otros? La nueva cirugía de bypass gástrico, que parece ser un gran avance para algunos casos de obesidad crónica causada por una tendencia obsesiva a comer, constituye sin duda una medida drástica, pero la opinión más extendida *hoy* en muchos círculos es que las personas con graves problemas de sobrepeso que se *resisten* a someterse a ella actúan de manera irresponsable (Gawand, 2001). Esto es algo que puede cambiar perfectamente a medida que vayamos aprendiendo más cosas sobre los efectos a largo plazo de la operación, tanto sobre las personas con tendencia obsesiva a comer como sobre la sociedad que les rodea y sus actitudes. Tales actitudes tienen un papel importante en el establecimiento de las condiciones en las que se toman las decisiones libres. Desórdenes relacionados con la comida como la bulimia y la anorexia nerviosa, por ejemplo, son mucho menos frecuentes entre las mujeres de los países musulmanes, donde el atractivo físico de las mujeres tiene un papel menos preeminente que en los países occidentales (Abed, 1998). Una revisión menor de las normas sociales, señala Gibbard, puede tener un profundo efecto sobre lo que piensan los individuos respecto a sus propias decisiones, y esto es una cuestión clave para distinguir las elecciones humanas de las elecciones de los animales.

Supongamos que tenemos una gran mancha morada en la espalda. Se trata sin duda de un rasgo biológico, pero no supone probablemente ningún rasgo especialmente relevante a nivel psicológico. Supongamos que tenemos en cambio una gran mancha morada en la nariz. Esto es una desgracia mucho mayor, más allá de que ambas decoloraciones puedan ser fisiológicamente inocuas, pues la mancha en la nariz interferirá sin duda seriamente en la imagen que tengamos de nosotros mismos, porque influirá en cómo nos vean y cómo nos traten los demás, y en cómo reaccionemos nosotros ante dicho tratamiento, y en cómo reaccionen ellos ante dichas reacciones, y así sucesivamente. Una nariz morada es un gran problema psicológico. El hecho de que sea un problema tan grande, sin embargo, es algo fácilmente reconocible por muchas personas, lo que puede llevar al establecimiento de políticas, prácticas y actitudes sociales que tiendan a minimizar sus efectos o, en todo caso, a mantenerlos controlados. Lo que antes era un rasgo biológico superficial de un organismo se convierte en un rasgo psicológico del mismo organismo, lo cual da origen a su vez a un rasgo político del mundo. Esta clase de cosas no acostumbran a ocu-

rrir en el mundo animal. Los etólogos de campo acostumbran a poner marcas en los animales que estudian para que les resulte más fácil reconocerlos a lo largo del tiempo. Miles de pájaros han vivido sus vidas con una cinta de color en una pata, y un número tal vez igual de mamíferos han seguido con sus asuntos con una chapa metálica numerada bien visible en sus orejas, y por el momento nadie ha dicho que estas marcas interfiriesen seriamente en sus vidas, ni en el sentido de aumentar ni en el de limitar sus oportunidades. Un ser humano que tuviera que aparecer en público con una chapa metálica en una oreja tendría que realizar importantes revisiones de sus proyectos y expectativas en la vida, por lo que la decisión de llevar una cosa así, sea propia o ajena, tiene necesariamente una dimensión política.

Esta sensibilidad a las repercusiones sociales y políticas que distingue la agencia humana de la agencia animal también sirve como base para fundar la responsabilidad humana en algo más prometedor que la indeterminación cuántica. Las negociaciones políticas de las que emergen nuestras prácticas actuales y nuestras presunciones sobre la responsabilidad no tienen nada que ver con el determinismo o el mecanicismo en general, sino más bien con una evaluación de la inevitabilidad —o la evitabilidad— de las características particulares de ciertos agentes y tipos de agentes. ¿Podemos enseñarles nuevos trucos a esos perros? Tal como vimos en el capítulo 3, hay un sentido no problemático en el que puede decirse que se ha producido un aumento de la competencia a lo largo del tiempo en un mundo determinista, así como una ampliación de las oportunidades y de lo que hacen con ellas los agentes deterministas particulares. Tal incremento de la competencia a lo largo del tiempo resulta completamente invisible si se adopta la estrecha noción de posibilidad implícita en la definición del determinismo: «En cada instante hay exactamente un único futuro posible». De acuerdo con esta noción, en un mundo determinista, en cualquier tiempo *t,* ningún ente *puede hacer* nada distinto de lo que está determinado a hacer en *t,* mientras que en un mundo indeterminista, en cualquier tiempo t, los entes pueden hacer tantas cosas distintas —al menos dos— como permita el tipo concreto de indeterminismo del que se trate, el cual será presumiblemente un profundo e inmutable hecho de la física que no se verá perturbado en lo más mínimo por los cambios en las prácticas, el conocimiento o la tecnología. El hecho evidente de que la gente hoy *puede hacer* más cosas de las que podía hacer antes se pierde de vista si concebimos de este modo la posibilidad y, sin embargo, se trata de un hecho que es tan importante como incuestionable.

En realidad, el problema al que se enfrentan actualmente los teóricos de todas las tendencias en el campo de la ética es la dificultad para manejar adecuadamente las implicaciones de *este* sentido específico del término «poder». Una de las escasas proposiciones incontrovertidas de la ética, que merece su propio y sencillo eslogan, es la de que «el *deber* implica el *poder»-,* sólo estamos obligados a hacer algo si somos capaces de hacerlo. Si somos honestamente incapaces de hacer X, entonces no es cierto que deberíamos hacer X. A veces se pretende que aquí reside el vínculo fundamental —y obvio— entre la libertad y la responsabilidad: puesto que sólo somos responsables de aquello que entra *en nuestras posibilidades,* y puesto que, si el determinismo es verdadero, sólo *podemos* hacer lo que sea que estemos determinados a hacer, nunca puede darse el caso de que *debamos* hacer otra cosa, pues nunca tenemos la posibilidad de hacer otra cosa. Pero, al mismo tiempo, es incluso más evidente que el crecimiento explosivo de las *posibilidades* que se ha producido en la historia reciente de la humanidad está dejando obsoletas muchas de nuestras nociones morales tradicionales sobre las obligaciones humanas, con independencia de cualquier consideración respecto al determinismo o al indeterminismo. El sentido de «poder» relevante desde el punto de vista moral no es el sentido de «poder» que depende del indeterminismo (si es que lo hay).

Supongamos que un adulto en plena posesión de sus facultades pero aquejado por alguna enfermedad nos pidiera ayuda para poner su cuerpo aún vivo en suspensión criogénica en espera de que en el futuro se descubriera una cura para su enfermedad, lo que es altamente improbable. ¿No sería eso asistencia ai suicidio? Hoy es razonable considerarlo así; mañana puede ser algo tan justificable como ayudar a que le administren anestesia a alguien para luego someterle a una operación quirúrgica que podría salvarle la vida. Nunca tuvimos que preocuparnos por la ética de la clonación, o por la posibilidad de ser sometidos a una vigilancia electrónica omnipresente, o por las drogas que pudieran tomar los adetas, o por la posibilidad de introducir mejoras genéticas en los embriones y nunca tuvimos que preocuparnos demasiado por la perspectiva de que pudieran desarrollarse refuerzos efectivos para la capacidad de los seres humanos de controlarse a sí mismos, pero a medida que van surgiendo tales innovaciones se nos plantea la necesidad de desarrollar una noción de responsabilidad que sea lo bastante sólida como para asimilarlas de manera no traumática.

«¡GRACIAS, LO NECESITABA!»

El cambio crucial de perspectiva que hará todo esto posible es una inversión que expone Stephen White en *The Unity of the Self* (1991, capítulo 8: «Moral Responsibility»). No tratemos de usar la metafísica para fundar la ética, propone White; hagámoslo al revés: usemos la ética para determinar cuál debe ser nuestro criterio «metafísico». En primer lugar, debemos mostrar qué tipo de justificación interna puede haber para que un agente acepte su propio castigo —diciendo: «¡Gracias, lo necesitaba!»— y luego debemos usar dicha idea para fijar y fundamentar una interpretación de nuestro lema central, *podría haber hecho otra cosa:* «Un agente podría haber hecho algo distinto de lo que hizo sólo en el caso de que esté justificada la adscripción de culpa y responsabilidad sobre dicho agente por la acción» (pág. 236). En otras palabras, el hecho de que la libertad sea algo valioso y deseable puede servir para fijar nuestra concepción de la libertad de un modo que no puede lograrse con mitos metafísicos. El argumento básico pretende cubrir todas las atribuciones de culpa o mérito moral, pero podemos simplificar el razonamiento si nos centramos en los casos de castigo impuesto por la autoridad («el Estado»), en representación de la clase más amplia de casos en los que, aunque no se haya cometido ningún *crimen,* un individuo culpa a otro de una mala acción. En muchos de los casos que entran en este conjunto más amplio no tiene por qué haber otro castigo previsto que el de ser reprendido (o ser objeto de resentimiento o malos pensamientos por parte del otro). Para controlar la generalidad del argumento podemos trasladarlo de vez en cuando del contexto legal (el Estado contra Jones) a un contexto moral (un padre que castiga a su hijo, por ejemplo).

El ideal para una institución de castigo, según White, sería que cada castigo estuviera justificado *a los ojos de la persona castigada.* Esto presupone que los agentes susceptibles de ser castigados sean lo bastante inteligentes, racionales e instruidos como para ser jueces competentes de la presunta justificación del castigo. Su (imaginaria) aquiescencia hacia su propio castigo sirve como referencia o elemento decisivo para determinar el umbral. Aquellos que son incapaces de realizar un juicio de este tipo no son probablemente lo bastante competentes como para disfrutar sin supervisión de las libertades que ofrece la ciudadanía, de modo que no les atribuimos responsabilidad (o no todavía, en el caso de los niños). Aquellos que son lo bastante competentes como para apreciar la justificación y aceptarla, son casos no problemáticos de malhechores culpables (eso di-

cen ellos mismos, y no tenemos razones plausibles para no aceptar su palabra). Eso deja sólo el caso de aquellos que son aparentemente competentes pero se resisten a dar su aquiescencia. Esos son los casos problemáticos, pero se encuentran presionados por ambos lados: por un lado, es presumible que deseen el estatus de ciudadano competente, dados los muchos beneficios que comporta, y, por el otro, temen el castigo, al que sólo pueden escapar declarándose —o revelándose— demasiado pequeños. (Si uno se hace lo bastante pequeño, puede externalizarlo prácticamente todo.) White señala, astutamente, que incluso un psicópata racional tendrá una justificación interna para apoyar las leyes que castigan a los psicópatas, porque le protegen de otros psicópatas y le confieren la libertad de perseguir sus propios intereses hasta donde le esté permitido.

Llegue o no a realizarse dicha ceremonia de la justificación, podemos imaginar el escenario donde tendría lugar. Supongamos que usted es el acusado. El Estado le dice: «Ha cometido una falta. Mala suerte, pero para el bien del Estado se le pide que acepte el castigo». Usted escucha los cargos, las pruebas, el veredicto. Supongamos que es usted culpable de los cargos que se le imputan. (Los equilibrios y los controles del sistema mantendrán la presión sobre el Estado para que resuelva debidamente los casos, y a usted se le animará a que utilice tal presunción en su defensa.) Pero ahora la cuestión es si usted es responsable del acto cometido. Podríamos plantear la cuestión en términos de: «¿Podría haber hecho otra cosa?», pero no buscaríamos el testimonio de metafísicos ni de físicos cuánticos. Buscaríamos pruebas *específicas* de su competencia, o circunstancias atenuantes. Consideremos, en particular, una defensa que alegue factores que estaban más allá de su control, factores que estaban allí desde mucho antes de su nacimiento, por ejemplo. Dichos factores sólo son relevantes si usted no sabía de su existencia. Si usted sabía que el suelo sobre el que estaba construyendo su casa había sido contaminado por residuos industriales un siglo antes, o *si debería haberlo sabido,* no puede alegarlo como un factor ajeno a su control. Pero ¿cómo pudo haberlo sabido? (El «deber» implica el «poder».) A medida que aumenta nuestra capacidad para adquirir conocimientos sobre los factores que tienen una influencia causal en nuestras acciones, nos volvemos cada vez más imputables por no conocer factores tanto externos (por ejemplo, el suelo contaminado) como internos (por ejemplo, su conocida obsesión por ganar dinero fácil: ¡debería haber hecho algo por resolverlo!). Una defensa del tipo: «No podía hacer otra cosa», que tal vez hubiera funcionado en otro tiempo, ya no es aceptable. Usted está obligado por las actitudes dominantes en la sociedad a estar al

corriente de los avances más recientes en todas las materias sobre las que usted pretenda ostentar alguna responsabilidad.

El Estado le invita a aceptar su castigo, a lo que, por supuesto, puede negarse, aunque si el Estado ha hecho bien su trabajo, debería aceptarlo. Es decir, el Estado puede ofrecerle una razón perfectamente defendible para ello. Si usted no la ve, es su problema. Si hay mucha gente que no la ve, es problema del Estado; ha establecido el umbral demasiado bajo o ha cometido algún otro error en el diseño de las leyes. ¿Cómo resolvemos los casos dudosos en un mundo que no es ideal sino real, donde hay personas que son incapaces de ver su justificación, o cuya aquiescencia es el resultado de la coerción o de un lavado de cerebro? La existencia de un conjunto no vacío de reos condenados que no son competentes para dar su aquiescencia al castigo es *inevitable*, pero no tiene por qué ser *inevitablemente grande*. De hecho, el sistema de umbrales negociados tiene la virtud de ser revisable a lo largo del tiempo para minimizar el conjunto de personas mal clasificadas. A medida que vamos desvelando errores de la justicia, los usamos como base para revisar nuestras políticas, y cuando descubrimos catego-rías de individuos que quedan por debajo del umbral de autocontrol defendido en el momento, nos enfrentamos a una cuestión política parecida a la de resolver si debemos cambiar las reglas para los permisos de conducir. Y si surgen nuevas tecnologías (cirugía, drogas, tratamientos, prótesis, sistemas educativos, sistemas de alarma, etc.) que pueden resultar efectivas para la mejora de las capacidades de aquellos que quedan por debajo del umbral, deberemos llegar a un compromiso coste-beneficios en función de si los efectos positivos son superiores a los negativos.

¿Pueden los pederastas actuar de otro modo? Algunos sí y algunos no, y deberíamos ver si se puede hacer algo para traspasar a cuantos sean posibles del segundo al primer grupo. Los que pueden actuar de otro modo son los que, *en caso* de reincidencia, insistirán en su *derecho* a ser castigados. Y no deberíamos prejuzgar la presunción de competencia que tienen a su favor para plantear una demanda de este tipo (aunque ésa será una cuestión que resolver en el juicio). Pero ¿acaso el mero hecho de la reincidencia, de cualquier reincidencia, no demuestra que después de todo *no* podían hacer otra cosa, al menos no en aquella situación concreta? No. Eso es un retorno injustificado a la noción estrecha del término «poder». Nuestras prácticas están asociadas a la noción más amplia del término, y ciertamente *imputamos* responsabilidad en tales individuos. Podrían haber hecho otra cosa en el sentido relevante. (Recordemos la versión trivial de este fenómeno presentada en el capítulo 3: el programa de ajedrez que no

se enrocaba pero que podría haberlo hecho, aunque operase en un mundo determinista, y, por lo tanto, nunca lo haría en esa circunstancia exacta.)

Sin embargo, ¿no es ésta una política demasiado arriesgada, teniendo en cuenta que es casi seguro que habrá algunos reincidentes? Tal vez sí, pero se trata de una cuestión política respecto a cuánto riesgo estamos dispuestos a asumir en nuestras vidas, no de una cuestión filosófica sobre si los pederastas son realmente libres o no, ni siquiera de una cuestión científica sobre qué es lo que hace que los pederastas se comporten de este modo. A medida que sepamos más cosas acerca de los factores —neuroquímicos, sociales, genéticos— que predisponen a la pederastía (y los cambiantes límites de la evitabilidad de dichos factores), podremos ciertamente reducir la incertidumbre, y por lo tanto el riesgo, de liberar a estas personas de su confinamiento, pero el riesgo siempre existirá. La cuestión política es cuánto riesgo estamos dispuestos a tolerar para conservar nuestra libertad como sociedad.

Durante siglos hemos vivido de acuerdo con la regla de que nadie podía ser castigado, ni detenido, por *su propensión a cometer un crimen,* pero durante todo este tiempo hemos sido muy conscientes de los riesgos que supone este admirable principio. ¿Qué hacemos con el ciudadano hasta ahora respetuoso con la ley que se acerca a su posible víctima con un arma peligrosa? ¿En qué momento debemos intervenir? ¿En qué punto pierde nuestro conciudadano su derecho a la no interferencia? ¿Tiene *derecho* a dar el primer golpe antes de que podamos emprender alguna acción contra él? A medida que vayamos descubriendo más cosas sobre las probabilidades y los factores subyacentes, habrá cada vez más presión para revisar nuestro admirable principio en interés de la seguridad pública. Nótese que disponemos ya de un gran número de inteligentes innovaciones legales que sirven a este propósito: preservan el admirable principio mediante la creación de nuevos crímenes que la gente comete como medio para cometer el crimen principal. Hemos promulgado una ley que prohibe a las personas llevar armas peligrosas en público, por ejemplo, o que instituye el nuevo crimen de conspiración para cometer otro crimen. En la actualidad es un crimen que personas con ciertos cuadros médicos oculten su solicitud de acceder a ciertos cargos de alto riesgo. Tenemos formas de traspasar la carga del conocimiento a los individuos, y ponerlos de este modo en posición de tomar decisiones análogas a la severa alternativa de la pederastía. Y, lo que es más importante, en la medida en que mantengamos el requisito de que dichas innovaciones deben pasar la prueba del «¡Gracias, lo necesitaba!», podemos conservar nuestra institu-

ción de la responsabilidad; podemos mantener a raya el espectro de la exculpación. Pregúntese a usted mismo: supongamos que *supiera* (merced a grandes avances de la ciencia) que su constitución le hace muy propenso a causar cierto tipo de daños en las personas a menos que se someta al tratamiento Z, que ayudaría a *evitar* una calamidad de este tipo; y supongamos que someterse a este tratamiento preservara su competencia en (prácticamente) todos los sentidos. ¿Estaría usted dispuesto a someterse al tratamiento? ¿Estaría a favor de una ley que convirtiera el hecho de someterse al tratamiento en una condición para que conservara su libertad? En otras palabras, ¿está usted seguro de que bajo esas condiciones usted tendría el *derecho* de dar el primer golpe? En el juicio, usted podría decir: «Tengo una enfermedad, señoría; estaba fuera de control. No podía hacer otra cosa», pero esto sería poco honesto por su parte si usted estaba al corriente de esta posibilidad. ¿Qué pasaría si este tratamiento debiera realizarse durante la infancia, antes de llegar a la edad de poder dar un consentimiento informado? ¿Estamos dispuestos a considerar la validez ética de dichas intervenciones preventivas? ¿Qué criterios de prueba deberíamos exigir antes de asumir una medida de «salud pública» de este tipo? (En la actualidad ya existen leyes que obligan a la vacunación, aunque sabemos lo suficiente como para tener una certeza moral de que algunos niños tendrán reacciones negativas a ella y morirán o quedarán discapacitados.) Cuanto más sabemos, más podemos hacer; cuanto más podemos hacer, más obligaciones tenemos. Tal vez terminemos por añorar los viejos tiempos en que la ignorancia era mejor excusa de lo que lo es hoy, pero no podemos pedirle al reloj que vuelva atrás.

Es el momento de recordar la desgracia del padre del capítulo 1, que carga con la responsabilidad —¿no es así?— de la muerte de su hijo. Presumiblemente todo el mundo controla las situaciones hasta donde es capaz de hacerlo, y, superado ese punto, simplemente deja de controlarlas. ¿Cómo puede ser justo exigir responsabilidades y castigar a esta persona, sólo porque alguna *otra* persona no habría cometido su error en las mismas circunstancias? ¿Acaso no es sólo una víctima de la mala suerte? Y ¿no es sólo buena suerte que nosotros no hayamos cedido a la tentación o que ninguna conspiración de circunstancias haya puesto al descubierto nuestras debilidades? Sí, la suerte es siempre un factor importante en nuestras vidas, pero, como somos conscientes de este hecho, tomamos las precauciones que nos parecen apropiadas para minimizar los efectos indeseables de la suerte, y luego asumimos la responsabilidad por lo que pueda ocurrir. Nótese que si este padre se hiciera lo bastante pequeño,

podría externalizar todo este episodio de su vida hasta convertirlo casi en una pesadilla, una cosa que le sucedió, no algo que hizo. Pero también tiene la opción de hacerse grande, y enfrentarse a la tarea mucho más ardua de construir un yo futuro que tenga esta terrible omisión en su biografía. Depende de él, pero cabría esperar que recibiera algo de ayuda de sus amigos. Esta es, sin duda, una buena ocasión para una Acción Autoformadora de las que habla Kane, y los seres humanos somos la única especie capaz de realizarlas, pero no hay ninguna necesidad de que sean indeterminadas.

¿SOMOS MÁS LIBRES DE LO QUE QUERRÍAMOS?

> *Tal vez si viéramos a dónde nos lleva el modelo supuestamente ideal de investigación, cambiaríamos de idea respecto a lo que es una investigación ideal. En cualquier caso, la única manera de que funcione un método como éste es mediante un trabajo lento y un gran esfuerzo en muchas direcciones distintas.*
>
> ALLAN GIBBARD, *Wise Choices, Apt feelings*

Nicholas Maxwell (1984) define la libertad como «la capacidad de conseguir lo valioso en una amplia variedad de circunstancias». Pienso que es una definición tan buena como cualquier otra. En particular, deja oportunamente abierta la cuestión de qué es lo valioso. Nuestra capacidad única de reconsiderar nuestras convicciones más profundas sobre lo que da valor a la vida nos obliga a tomarnos en serio el descubrimiento de que no hay ningún límite perceptible en lo que podemos considerar valioso. Sólo depende de nosotros. Ésta es una perspectiva temible para algunas personas, pues les parece que abre las puertas al nihilismo y al relativismo, a la anarquía y al abandono de los mandamientos divinos. *¡Detengan a ese cuervo!*

Pienso que estas personas deberían tener más fe en los demás seres humanos y apreciar lo tremendamente hábiles y perspicaces que son, lo bien que les ha dotado la naturaleza y la cultura para diseñar adecuadamente ordenamientos sociales que maximizan la libertad de todos y para participar en ellos. Lejos de ser anárquicos, dichos ordenamientos están —y deben estar— exquisitamente concertados para generar un equilibrio entre la seguridad y la flexibilidad. Si no podemos alcanzar la universalidad (la chovinista palabra que usa el *Homo sapiens* para referirse a la aceptación en el conjunto de la especie), podemos aspirar al menos a lo

que Alian Gibbard llama «el tribalismo de la tribu más amplia» (Gibbard, 1990, pág. 315). Pero tal vez seamos capaces de alcanzar la verdadera universalidad. Lo hemos hecho en otros dominios. El problema de los filósofos es negociar la transición del «ser» al «deber ser» o, más precisamente, mostrar cómo podemos ir más allá del hecho «meramente histórico» de que ciertas costumbres y políticas han obtenido una amplia aceptación social, y convertirlas en normas que exijan el asentimiento de todos los seres racionales. Se conocen casos en los que se ha realizado esta maniobra con éxito. El aupamiento ha funcionado en el pasado, y puede funcionar ahora también. No necesitamos ningún gancho colgado del cielo.

Consideremos el curioso problema que supone dibujar una línea recta. Una línea *realmente* recta. ¿Cómo podemos hacerlo? Usando una regla, claro. ¿Y dónde la conseguimos? A lo largo de los siglos hemos ido refinando nuestras técnicas para obtener reglas cada vez más rectas, usando unas para supervisar y corregir las otras en un proceso que no ha dejado de elevar el umbral de la precisión. En la actualidad disponemos de máquinas con una precisión de una millonésima de pulgada en toda su longitud, y desde nuestra privilegiada posición actual no estamos lejos del ideal normativo prácticamente inalcanzable pero fácilmente concebible de una regla *realmente* recta. Descubrimos dicho ideal normativo, la eterna forma platónica de la recta, si se quiere, gracias a nuestra actividad creativa. También descubrimos la aritmética, y otros tantos sistemas de verdades absolutas e intemporales. Tal como dice Gibbard, *tal vez* no encontremos un límite parecido en nuestra búsqueda de un sistema ético, pero no veo razón alguna *a priori* para descartar dicho horizonte, desde el momento en que ya hemos desarrollado el ideal de una sociedad libre en la que pueda desarrollarse una investigación libre. El carácter normativo implícito en dichos descubrimientos humanos —¿o son más bien inventos?— es en sí mismo uno de los frutos de los-procesos evolutivos, tanto genéticos como culturales, que han hecho que seamos lo que somos, como resultado del aprovechamiento y la amplificación de miles de millones de colisiones fortuitas, los «accidentes congelados» de la historia, tal como los ha llamado Francis Crick, que nos han llevado hasta nuestro estado actual. El proceso colectivo de ingeniería memética que llevamos a cabo desde hace miles de años sigue adelante, y este libro forma parte de este proceso. No descubre ningún punto arquimédico desde el que mover el mundo, pero tal vez pueda contribuir al refinamiento de la comprensión que tenemos de nosotros mismos y de nuestras circunstancias.

La libertad de *pensamiento* y *acción* necesaria para descubrir la verdad es precursora, tal como hemos visto, del ideal más expansivo de la libertad política o civil, un meme que parece difundirse con facilidad. Gracias a Dios, es mucho más infeccioso que el fanatismo. El secreto está en la calle. No hay manera de que la ignorancia forzada pueda ganar la partida a largo plazo. No es fácil desinformar a la gente. A medida que las tecnologías de la comunicación hacen cada vez más difícil que los líderes aislen a su pueblo de toda información exterior, y a medida que las realidades económicas del siglo **XXI** dejan cada vez más claro que la educación es la inversión más importante que puede hacer un padre en beneficio de un hijo, el proceso va a hacerse imparable en todo el mundo y sus efectos serán masivos. Todo el cajón de sastre de la cultura popular, toda la basura y la escoria que se acumula en los rincones de una sociedad libre, inundará esas regiones relativamente prístinas al tiempo que lo hacen los tesoros de la educación moderna, la igualdad de derechos para las mujeres, las mejoras en la salud, los derechos de los trabajadores, los ideales democráticos y la apertura a las culturas ajenas. Tal como demuestra claramente la experiencia de la antigua Unión Soviética, los peores rasgos del capitalismo y la tecnología son algunos de los replicadores más robustos de esta explosión demográfica de memes, y no faltarán los motivos para la xenofobia, el ludismo y la tentadora «higiene» del fundamentalismo retrógrado.

Tal como demuestra Jared Diamond en *Armas, gérmenes y acero* (1997), fueron los gérmenes europeos los que llevaron casi a la extinción a las poblaciones del hemisferio occidental, pues la historia de dichos pueblos no les había permitido desarrollar tolerancia a los mismos. En el próximo siglo serán nuestros memes, tanto los tónicos como los tóxicos, los que harán estragos en aquellas partes del mundo que todavía no estén preparadas. No podemos presumir que los demás tengan *nuestra* capacidad para tolerar los excesos tóxicos de la libertad, ni podemos exportar esta capacidad simplemente como un producto más. La educabilidad prácticamente ilimitada de cualquier ser humano nos da, sin embargo, la esperanza del éxito, pero diseñar e implementar los mecanismos de vacunación cultural necesarios para evitar el desastre, al tiempo que respetamos los derechos de aquellos que más necesitan dicha vacuna, será una tarea urgente y compleja, que no requerirá únicamente una mejor ciencia social, sino también sensibilidad, imaginación y coraje. El principal campo de batalla de este siglo será el de la salud pública, cuyo concepto deberá ampliarse para incluir la salud cultural.[1]

1. Los dos párrafos precedentes están tomados de Dennett, 1999b.

LA LIBERTAD HUMANA ES FRÁGIL

Las ballenas vagan por el océano, los pájaros vuelan ligeros por encima de nuestras cabezas y, según un viejo chiste, un gorila de más de 200 kilos se sienta donde le da la gana, pero ninguna de estas criaturas es libre en el sentido en que pueden serlo los seres humanos. La libertad humana no es una ilusión; es un fenómeno objetivo, distinto de todas las demás condiciones biológicas y que sólo se encuentra en una especie, la nuestra. Las diferencias entre los agentes humanos autónomos y los demás agregados naturales son visibles no sólo desde una perspectiva antropocéntrica, sino también desde los más objetivos de los puntos de vista alcanzables (el plural es importante). La libertad humana es real —tan real como el lenguaje, la música y el dinero—, de modo que puede ser estudiada desde un punto de vista serio, objetivo y científico. Pero igual que el lenguaje, la música, el dinero y otros productos de la sociedad, su persistencia se ve afectada por lo que creemos sobre ella. No es ninguna sorpresa, pues, que nuestros intentos de estudiarla desapasionadamente se vean distorsionados por el miedo de matar torpemente el espécimen que tenemos bajo el microscopio.

La libertad humana es más joven que la especie. Sus caracteres principales tienen únicamente unos miles de años de antigüedad —un parpadeo dentro de la historia evolutiva—, pero en ese tiempo tan breve ha transformado el planeta de una forma tan palpable como pudieran hacerlo grandes transiciones biológicas como la creación de una atmósfera rica en oxígeno y la creación de la vida multicelular. La libertad tuvo que evolucionar igual que todos los demás elementos de la biosfera, y continúa su evolución en la actualidad. La libertad es real hoy en algunas partes afortunadas del planeta, y aquellos que la aman tienen razón de hacerlo, pero está lejos de ser inevitable, y lejos de ser universal. Si llegamos a comprender mejor su origen, tal vez podamos orientar mejor nuestros esfuerzos para preservarla de cara al futuro, y protegerla de sus muchos enemigos naturales.

Nuestros cerebros han sido diseñados por la selección natural, y todos los productos de nuestros cerebros han sido diseñados del mismo modo, aunque en una escala temporal mucho más reducida, por procesos físicos en los que no puede discernirse ninguna exención de los principios causales. ¿Cómo es posible, entonces, que nuestros inventos, nuestras decisiones, nuestros pecados y nuestros éxitos sean distintos de las bellas pero amorales telas que tejen las arañas? ¿Qué diferencia hay, desde un punto de vista moral, entre una tarta de manzana preparada con todo el cariño por alguien como regalo de reconciliación, y una manzana diseñada

«inteligentemente» por la evolución para atraer a un frugívoro, el cual se encargará de repartir sus semillas a cambio de algo de fructosa? Si consideramos que eso son preguntas retóricas, en el sentido de que sólo un milagro podría distinguir nuestras creaciones de las ciegas e inconscientes creaciones de un mecanismo material, no dejaremos de dar vueltas a los problemas tradicionales de la libertad y el determinismo, en una espiral de misterio e incomprensión. Los actos humanos —los actos de amor y de genio, así como los crímenes y los pecados— están simplemente demasiado lejos de los movimientos de los átomos, sean aleatorios o no, como para que podamos descubrir la manera de integrarlos en un único esquema coherente. Los filósofos han intentado durante milenios salvar esta brecha con un par de maniobras atrevidas, sea a base de poner a la ciencia o al orgullo humano en su lugar, o bien declarando (correctamente, pero de manera poco persuasiva) que la incompatibilidad es sólo aparente, sin entrar en los detalles. Cuando tratamos de responder a estas preguntas, de esbozar las vías no milagrosas que puedan llevarnos desde los ciegos átomos hasta las acciones libremente escogidas, abrimos nuevos espacios a la imaginación. La compatibilidad entre la libertad y la ciencia (determinista o indeterminista, no importa) no es tan inconcebible como podría parecer.

Los temas tratados en este libro no son sólo problemas académicos, seductores acertijos conceptuales aún por resolver o fenómenos curiosos todavía no integrados en una buena teoría. Muchas personas los ven como cuestiones de vida o muerte, y eso los convierte en cuestiones de vida o muerte, pues los miedos de las personas tienden a amplificar las implicaciones previstas por los diferentes análisis y a distorsionar los argumentos, hasta convertirlos en meros instrumentos de propaganda, para bien o para mal. La resonancia emocional de la palabra «libertad», como la de la palabra «Dios», garantiza una audiencia partidista, dispuesta a saltar sobre cualquier paso en falso, cualquier amenaza, cualquier concesión. El resultado es que la tradición acostumbra a tener carta blanca, o casi. A manera de estrategia práctica, la mayoría de la gente parece inclinarse a pensar que las doctrinas suscritas por la tradición deberían pasar sin examen alguno, en la medida de lo posible, y cuestionarlas es ciertamente como tocar un avispero. Y así es como sobrevive el pensamiento tradicional, en gran medida incuestionado, y con los años no hace más que acumular nuevas capas de invulnerabilidad injustificada.

He tratado de mostrar, con la ayuda de muchos otros pensadores, que podemos y debemos sustituir estas sacrosantas pero endebles tradiciones por unos fundamentos más naturalistas. Da miedo abandonar unos pre-

ceptos tan venerables como el imaginario conflicto entre el determinismo y la libertad, y la falsa seguridad que da pensar que la cadena termina en un Yo o un Alma milagrosa. El análisis filosófico, sin embargo, no es suficiente por sí solo para animar un cambio tan drástico en nuestra manera de pensar, por más que sea correcto en lo fundamental, y tal vez el aspecto más radical de este libro escrito por un filósofo sea la preeminencia que otorga a la obra de no-filósofos. Mi idea es que los filósofos, *en cuanto filósofos,* no pueden pretender estar haciendo de manera profesional su trabajo en relación con los temas que les son más propios a menos que presten atención al pensamiento de psicólogos como Daniel Wegner y George Ainslie, de economistas como Robert Frank, de biólogos como Richard Dawkins, Jared Diamond, Edward O. Wilson, David Sloan Wilson, y de otros cuyas ideas han ocupado un lugar destacado en este libro. Por supuesto, no soy el único filósofo que defiende esta perspectiva. Excelentes filósofos como Jon Elster, Alian Gibbard, Philip Kitcher, Alexander Rosenberg, Don Ross, Brian Skyrms, Kim Sterelny y Elliott Sober han llegado más lejos que yo en la exploración de estas ricas vetas de material filosófico, en el proceso de clarificar tanto la ciencia como la filosofía.

No sólo he prodigado mi atención a las ideas de los no-filósofos; en el proceso he ignorado las ideas de unos cuantos filósofos de gran reputación, he pasado por alto sin apenas una mención algunas controversias que son objeto de un vivo debate en mi disciplina. Debo una explicación a los participantes en tales debates. ¿Dónde están, podrían preguntarse, mis refutaciones, mis pruebas, mis argumentos filosóficos que demuestran la invalidez de sus elaborados análisis? He aportado unos cuantos: el tiro al hoyo de Austin, la facultad de razonamiento práctico de Kane y la autonomía de Mele, por ejemplo, han recibido por mi parte la clase de atención detallada que esperan los filósofos. Respecto a los demás, he decidido traspasarles a ellos la carga de la prueba. Se requieren unos cuantos supuestos de fondo comunes para establecer una controversia filosófica, y, aunque no lo he demostrado, me he convencido a mí mismo de que mis relatos y observaciones informales cuestionan algunas de sus premisas fundamentales, lo que convierte sus controversias en opcionales para mí, por más entretenidas que resulten para aquellos que están enzarzados en ellas. Podría haber dicho exactamente cómo y por qué, pero eso hubiera requerido cien páginas o más de densa argumentación y exégesis textual, para terminar con un veredicto de falsa alarma, un anticlímax que prefiero rehuir en lo posible. Es sin duda una decisión arriesgada por mi parte, pues les da la posibilidad de demostrar que he cometido el error de sub-

estimar la inevitabilidad de las presuposiciones que comparten, pero es un riesgo que estoy dispuesto a asumir.

Mi intención en este libro ha sido demostrar que si aceptamos la «extraña inversión del razonamiento» de Darwin, podemos reconstruir los mejores y más profundos pensamientos humanos sobre moral, sentido, ética y libertad. Lejos de ser enemiga de dichos conceptos tradicionales, la perspectiva evolutiva es un aliado indispensable de los mismos. No pretendo reemplazar el abundante trabajo realizado hasta el momento en el campo de la ética por una *alternativa* darwinista, sino más bien asentar dicho trabajo sobre los cimientos que merece: una visión realista, naturalista, potencialmente unificada del lugar que ocupamos en la naturaleza. Reconocer nuestro carácter único 'como animales reflexivos y capaces de comunicarse no requiere ningún «excepcionalismo» humano que levante un puño desafiante frente a Darwin y descarte cualquier intuición procedente de un sistema de pensamiento magníficamente articulado y empíricamente contrastado. Podemos comprender por qué nuestra libertad es mayor que la de las demás criaturas, y en qué medida esta superior capacidad trae consigo implicaciones morales: *noblesse oblige.* Estamos en una posición privilegiada para decidir lo que haremos a continuación, porque disponemos del más amplio conocimiento posible y, por lo tanto, de la mejor perspectiva sobre el futuro. Lo que el futuro depara a nuestro planeta depende de todos nosotros, de nuestra reflexión conjunta.

NOTAS SOBRE FUENTES Y LECTURAS COMPLEMENTARIAS

La antología de Robert Kane, *The Oxford Handbook of Free Will* (2001), recoge artículos de nuevo encargo realizados por los principales autores de literatura filosófica en los últimos años, y los lectores encontrarán útiles análisis de los temas tratados en este libro.

Las complejas cuestiones relativas al castigo y la reincidencia reciben un buen tratamiento en Quinsey y otros, *Violent Offenders: Appraising and Managing Risk* (1998), un estudio amplio e informado sobre la predicción y el tratamiento de estos casos, que dedica especial atención a los psicópatas. Uno de sus descubrimientos más sorprendentes es que los psicópatas que siguen un programa de reeducación en sensibilidad social y relaciones interpersonales durante su encarcelamiento son más propensos a cometer crímenes violentos al ser liberados: «Especulamos, pues, con la posibilidad de que los pacientes aprendieran mucho en el programa intensivo, pero que aplicaran sus nue-

vas habilidades a usos muy distintos de los pretendidos» (pág. 89). Los filósofos deben reconsiderar los supuestos de fondo —las simplificaciones— que acostumbran a invocar cuando tratan el tema de los psicópatas y otros reos problemáticos. Como sucede siempre, la imaginación del filósofo dejada a su aire, sin contacto con los hechos, es un instrumento demasiado burdo para tratar un conjunto de problemas tan delicados e importantes.

The Unity ofthe Self (1991), de Stephen White, sobre todo los capítulos 8 y 9, contiene un penetrante y detallado análisis de algunas de las cuestiones que planteo aquí a grandes trazos, y desarrolla argumentos que deberían satisfacer a los escépticos, sobre todo en relación con la necesidad y la validez de la inversión que propone. En particular, recomiendo su análisis de las limitaciones de los esfuerzos filosóficos previos sobre estas cuestiones.

Foundations of Mechanical Accuracy (1970), de Wayne Moore, es un libro fascinante sobre la historia del proceso de aupamiento que ha dado lugar a los actuales (bueno, de la década de 1970) estándares de rectitud y precisión.

Algunos lectores de este libro han echado en falta un tratamiento de la creatividad y la autoría humanas. Éste fue el tema de mi discurso presidencial ante la División Este de la Asociación Americana de Filosofía, en diciembre de 2000 (Dennett, 2000b).

La relación entre la libertad personal y la libertad política recibe un incisivo tratamiento de Philip Pettit en *A Theory of Freedom: From the Psychology to the Politics ofAgency* (2001) y de Robert Nozick en el último capítulo de *Invariances,* «The Genealogy of Ethics» (2001). La importancia de la cultura, sobre todo de la organización política y económica, en el mantenimiento y promoción de la libertad queda demostrada en *Desarrollo y libertad* (1999), de Amartya Sen.

Mientras daba los últimos toques al libro recibí una copia por correo electrónico del nuevo libro de Merlin Donald, *A MindSo Rare: The Evolution of Human Consciousness* (2001). Donald deja claro desde la primera página que concibe su libro como una especie de antídoto contra dos míos: *La conciencia explicada* (1991a) y *La peligrosa idea de Darwin* (1995). Sin embargo, el último capítulo de la obra de Donald, «The Triumph of Consciousness», podría servir perfectamente como último capítulo de este libro. ¿Cómo es esto posible? Porque Donald, igual que muchos otros, ha subestimado enormemente lo que puede aportarnos la «extraña inversión del razonamiento» de Darwin. Dice en su Prólogo: «Este libro propone que la mente humana es distinta que cualquier otra mente que pueda haber en el planeta, no a causa de sus caracteres biológicos, que no son cualitativamente únicos, sino por su capacidad de generar y asimilar la cultura» (pág. xiii). Exactamente.

Bibliografía

Abed, Riadh, «The Sexual Competition Hypothesis for Eating Disorders», *British Journal of Medical Psychology,* vol. 17, n°4, 1998, págs. 525-547.

Ainslie, George, *Breakdown of Will,* Cambridge, Cambridge University Press, 2001.

Akins, Kathleen, «A Question of Content», en Andrew Brook y Don Ross (comps.), *Daniel Dennett,* Cambridge, Cambridge University Press, 2002, págs. 206-246.

Allison, Henry A., «We Can Act Only under the Idea of Freedom», *Proceedings of the American Philosophical Association,* vol. 71, n° 2,1997, págs. 39-50.

Astington, Janet, P. L. Harris y D. R. E. Olson (comps.), *Developing Theories of Mind,* Nueva York, Cambridge University Press, 1988.

Aunger, Robert (comp.), *Darwinizing Culture: The Status of Memetics as a Science,* Oxford, Oxford University Press, 2000.

—, *The Electric Meme: A New Theory of How We Think and Communicate,* Nueva York, Free Press, 2002 (trad, cast.: *El meme eléctrico,* Barcelona, Paidós, 2004).

Austin, John, «Ifs and Cans», en J. O. Urmson y G. Warnock (comps.), *Philosophical Papers,* Oxford, Clarendon Press, 1961.

Avital, Eytan y Eva Jablonka, *Animal Traditions: Behavioral Inheritance in Evolution,* Cambridge, Cambridge University Press, 2000.

Baker, Nicholson, *The Size of Thoughts: Essays and Other Lumber,* Nueva York, Random House, 1996.

Baron-Cohen, Simon, *Mindblindness: An Essay on Autism and Theory of Mind,* Cambridge, MA, MIT Press, 1995.

Baron-Cohen, Simon, H. Tager-Flusberg y D. Cohen (comps.), *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience,* Oxford, Oxford University Press, 2000.

Belie, Michael, *Darwin's Black Box: The Biochemical Challenge to Evolution,* Nueva York, Free Press, 1996.

Berry, Michael, «Regular and Irregular Motion», American Institute of Physics, accesible en http://www.phy.bris.ac.uk/staff/berry-mv.html, 1978.

Bingham, Paul M., «Human Uniqueness: A General Theory», *Quarterly Review of Biology,* n° 74,**1999,** págs. **133-169.**

Binmore, K. G., *Game Theory and the Social Contract*, vol. 2, *Just Playing,* Cambridge, MA, MIT Press, 1998.

Blackmore, Susan, *The Meme Machine,* Oxford, Oxford University Press, 1999 (trad, cast.: *La máquina de los memes,* Barcelona, Paidós, 2000).

Boone, James L. y Eric Alden Smith, «A Critique of Evolutionary Archaeology», *Current Anthropology,* n° 39, suplemento de 1998, págs. 104-151.

Boyd, R. y P. Richerson, «Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups», *Ethology and Sociobiohcgy,* n° 13, 1992, págs. 171-195.

Boyer, Pascal, *Religion Explained: The Evolutionary Origins of Religious Thought,* Nueva York, Basic Books, 2001.

Burkert, Walter, *Creation of the Sacred: Tracks of Biology in Early Religions,* Cambridge, MA, Harvard University Press, 1996.

Cabell, James Branch (1919), *Beyond Life: Dizain des Démiurges,* R. M. McBride, reimpr. 1929.

Calvin, William, *The Cerebral Symphony: Seashore Reflections on the Structure of Consciousness,* Nueva York, Bantam, 1989.

Campbell, Donald, «On the Conflicts Between Biological and Social Evolution and Between Psychology and Moral Tradition», *American Psychologist*, diciembre de 1975, págs. 1.103-1.126.

Cartmill, Matt, *A View to a Death in the Morning: Hunting and Nature through History,* Cambridge, MA, Harvard University Press, 1993.

Chisholm, Roderick, «Human Freedom and the Self», The Lindley Lecture, University of Kansas, 1964; reimpreso en Gary Watson (comp.), *Free Will,* Oxford, Oxford University Press, 1982.

Churchland, Patricia S., «On the Alleged Backwards Referral of Experiences and Its Relevance to the Mind-Body Problem», *Philosophy of Science,* n° 48,1981, págs. 165-181.

Churchland, Paul, *The Engine of Reason, The Seat of the Soul,* Cambridge, MA, MIT Press, 1995.

Clark, Thomas, «Review of *The Volitional Brain»,* en Libet y otros, 1999, págs. 271-285.

Cloak, F. T., «Is a Cultural Ethology Possible?», *Human Ecology,* n° 3, 1975, págs. 161-182.

Coleman, Mary, «Decisions in Action: Reasons, Motivation, and the Connection Between Them», tesis doctoral, Harvard University, Philosophy Department, 2001.

Cronin, Helena, *The Ant and the Peacock: Altruism and Sexual Selection from Darwin to Today,* Cambridge, Cambridge University Press, 1991.

Darwin, Charles, *On the Origin of Species by Means of Natural Selection,* Lon-

dres, Murray (edición facsímil de Harvard University Press), 1859 (trad, cast.: *El origen de las especies*, Madrid, Alianza, 2003).

Dawkins, Richard (1976), *The Selfish Gene,* 2ª ed., Oxford, Oxford University Press, 1989 (trad, cast.: *El gen egoísta,* Barcelona, Salvat, 2000).

—, *The Extended Phenotype: The Gene as the Unit of Selection,* San Francisco, Freeman, 1982.

—, «Viruses of the Mind», en Bo Dahlbom (comp.), *Dennett and his Critics,* Oxford, Blackwell, 1993.

—, *Climbing Mount Improbable,* Nueva York, Norton, 1996 (trad, cast.: *Escalando el monte Improbable,* Barcelona, Tusquets, 1998).

De Waal, Frans B. M., *Good Natured: The Origins of Right and Wrong in Humans and Other Animals,* Cambridge, MA, Harvard University Press, 1996 (trad, cast.: *Bien natural: los orígenes del bien y del mal en los humanos y otros animales,* Barcelona, Herder, 1997).

Dennett, Daniel C., «Why the Law of Effect Will Not Go Away», *Journal for the Theory of Social Behaviour,* n" *5,* 1975, págs. 169-187.

—, *Brainstorms: Philosophical Essays on Mind and Psychology,* Montgomery, VT, Bradford Books, 1978.

—, *Elbow Room: The Varieties of Free Will Worth Wanting,* Cambridge, MA, MIT Press and Oxford University Press, 1984.

—, *The Intentional Stance,* Cambridge, MA, MIT Press, 1987 (trad, cast.: *La actitud intencional,* Barcelona, Gedisa, 1991).

—, «The Interpretation of Texts, People, and Other Artifacts», *Philosophy and PhenomenologicalResearch,* n° 50, 1990, págs. 177-194.

—, *Consciousness Explained,* Boston, Little, Brown, 1991a (trad, cast.: *La conciencia explicada,* Barcelona, Paidós, 1995).

—, «Real Patterns», *Journal of Philosophy,* n° 88, 1991b, págs. 27-51; reimpreso en *Brainchildren.*

—, «Learning and Labeling» (comentario sobre A. Clark y A. Karmiloff-Smith, «The Cognizer's Innards»), *Mind and Language,* vol. 8, n° 4, 1993, págs. 540-547.

—, *Darwin's Dangerous Idea: Evolution and the Meanings of Life,* Nueva York, Simon and Schuster, 1995 (trad, cast.: *La peligrosa idea de Darwin: evolución y significados de la vida,* Barcelona, Galaxia Gutenberg, 2000).

—, *Kinds of Minds:Toward an Understanding of Consciousness,* Nueva York, Basic Books, 1996a (trad, cast.: *Tipos de mente: hacia una comprensión de la conciencia,* Madrid, Debate, 2000).

—, «Cow-sharks, Magnets, and Swampman», *Mind and Language,* vol. 11, n° 1, 1996b, págs. 76-77.

—, «Producing Future by Telling Stories», en K. Ford y Z. Pylyshyn (comps.), *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence,* Norwood, NJ, Ablex, 1996c, págs. 1-7.

, «Appraising Grace: What Evolutionary Good Is God?» (reseña de Walter Burkert, *Creation of the Sacred: Tracks of Biology in Early Religions), The Sciences,* enero-febrero de 1997a, págs. 39-44. (Una versión más extensa, titulada «The Evolution of Religious Memes: Who or What-Benefits?», con una respuesta de Walter Burkert aparece en *Method and Theory in the Study of Religion,* n° 10, 1998, págs. 115-128).

, «How to Do Other Things with Words», Royal Institute Conference on Philosophy of Language, en John Preston (comp.), *Philosophy,* n° 42, suplemento, Cambridge University Press, 1997b, págs. 219-235.

, «The Case of the Tell-Tale Traces: A Mystery Solved; a Skyhook Grounded», 1997C, http://ase.tufts.edu/cogstud/papers/behe.htm

, *Brainchildren: Essays on Designing Minds,* Cambridge, MA, MIT Press, 1998a.

, comentario sobre Boone y Smith, «A Critique of Evolutionary Archaeology», *Current Anthropology,* n°39, suplemento, 1998b, págs. 157-158.

, reseña de John Haugeland, *Having Thought: Essays in the Metaphysics of Mind,Journal of Philosophy,* n° 96, 1999a, págs. 430-435.

, «Protecting Public Health», en «Predictions: 30 Great Minds on the Future», *Times Higher Education Supplement,* marzo de 1999b, págs. 74-75.

, «Making Tools for Thinking», en Dan Sperber (comp.), *Metarepresentations: A Multidisciplinary Perspective,* Oxford, Oxford University Press, 2000a.

, «In Darwin's Wake, Where Am I?» (2000B), discurso presidencial de la American Philosophical Association Eastern Division, *Proceedings and Addresses of the American Philosophical Association,* n° 75, noviembre de 2001, págs. 13-30, accesible en http://ase.tufts.edu/cogstud

, «Collision Detection, Muselot, and Scribble: Some Reflections on Creativity», en David Cope (comp.), *Virtual Music,* Cambridge, MA, MIT Press, 2001a.

, «The Evolution of Culture», *The Monist,* vol. 84, n° 3,2001b, págs. 305-324.

, «The Evolution of Evaluators», en Antonio Nicita y Ugo Pagano (comps.), *The Evolution of Economic Diversity,* Londres, Routledge, 2001c.

, «The New Replicators», en M. Pageis (comp.), *Encyclopedia of Evolution,* Oxford, Oxford University Press, 2002a.

, «The Baldwin Effect: A Crane, not a Skyhook», en Bruce Weber and David Depew (comps.), *Evolution and Learning: The Baldwin Effect Reconsidered,* Cambridge, MA, MIT Press, 2002b.

, «Altruists, Chumps, and Inconstant Pluralists» (comentario sobre Sober y Wilson, 1998), *Philosophy and Phenomenological Research,* en preparación a.

, reseña de Avital y Jablonka, 2000, *Journal of Evolutionary Biology,* en preparación b.

, «From Typo to Thinko», en Steven Levinson (comp.), *Evolution and Culture,* Cambridge, MA, MIT Press, en preparación c.

—,*The Science of Consciousness: Removing the Philosophical Obstacles,* Jean Nicod Lectures, discurso pronunciado en París en noviembre de 2001, Cambridge, MA, MIT Press, en preparación d.

Dennett, Daniel C. y Marcel Kinsbourne, «Time and the Observer: The Where and When of Consciousness in the Brain», *Behavioral and Brain Sciences,* n° 15, 1991, págs. 183-247.

Densmore, Shannon y Daniel Dennett, «The Virtues of Virtual Machines», *Philosophy and PhenomenologicalResearch,* n° 59,1999, págs. 747-767.

Diamond, Jared, *Guns, Germs, and Steel: The Fates of Human Societies,* Nueva York, Norton, 1997 (trad, cast.: *Armas, gérmenes y acero,* Madrid, Debate, 1998).

Dickerson, Debra J., *An American Story,* Nueva York, Pantheon, 2000.

Donald, Merlin, *A Mind So Rare: The Evolution of Human Consciousness,* Nueva York, Norton, 2001.

Dooling, Richard, *Brain Storm,* Nueva York, Random House, 1998.

Drescher, Gary, *Made-Up Minds: A Constructivist Approach to Artificial Intelligence,* Cambridge, MA, MIT Press, 1991.

Fischer, John Martin y Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility,* Nueva York, Cambridge University Press, 1998.

Frank, Robert H., *Passions within Reason: The Strategic Role of the Emotions,* Nueva York, Norton, 1988.

Frankfurt, Harry, «Alternative Possibilities and Moral Responsibility», *journal of Philosophy,* n° 65,1969, págs. 829-833.

—, «Freedom of the Will and the Concept of a Person», *journal of Philosophy,* n° 68,1971, págs. 5-20.

Frayn, Michael, *Headlong,* Londres, Faber and Faber, 1999.

French, Robert M., *The Subtlety of Sameness: A Theory and Computer Model of Analogy-Making,* Cambridge, MA, MIT Press, 1995.

Gallagher, Shaun, «The Neuronal Platonist», en conversación con Michael Gazzaniga, *journal of Consciousness Studies,* vol. 5, n° 5-6, 1998, págs. 706-717.

Gawand, Atul, «The Man Who Couldn't Stop Eating», *The New Yorker,* 9 de julio de 2001, págs. 66-75.

Gazzaniga, Michael, *The Mind's Past,* Berkeley, University of California Press, 1998 (trad, cast.: *El pasado de la mente,* Barcelona, Andrés Bello, 1999).

Gibbard, Allan, *Wise Choices, Apt Feelings: A Theory of Normative judgment,* Cambridge, MA, Harvard University Press, 1990.

Giorelli, Giulio, «Si, abbiamo un anima. Ma é fatta di tanti piccoh robot», entrevista con Daniel C. Dennett, *Corriere della Sera,* Milán, 28 de abril de 1997.

Goffman, Erving, *The Presentation of Self in Everyday Life,* Nueva York, Anchor Doubleday, 1959 (trad, cast.: *La presentación de la persona en la vida cotidiana,* Madrid, H. F. de Murguía, 1987).

Goldschmidt, Tijs, *Darwin's Dreampond,* Cambridge, MA, MIT Press, 1996.
Gopnik, Adam, «Culture Vultures», *The New Yorker,* 24 de mayo de 1999, págs. 27-28.
Gould, Stephen Jay, *Ever Since Darwin,* Nueva York, Norton, 1978.
Gray, Russell D. y F. M. Jordan, «Language Trees Support the Express-train Sequence of Austronesian Expansion», *Nature,* n° 405,2000, págs. 1.052-1.055.
Greenough, W. T. y F. R. Volkmar, «Rearing Complexity Affects Branching of Dendrites in the Visual Cortex of the Rat», *Science,* n° 176,1972, págs. 1.445-1.447.
Haig, David, «Genomic Imprinting and the Theory of Parent-Offspring Conflict», *Developmental Biology,* n° 3,1992, págs. 153-160.
—, *Genomic Imprinting and Kinship,* New Brunswick, NJ, Rutgers University Press, 2002.
Haig, David y A. Grafen, «Genetic Scrambling as a Defence against Meiotic Drive», *Journal of Theoretical Biology,* n° 153, 1991, págs. 531-558.
Hamilton, William D., *Narrow Roads of Gene Land,* vol. 1, *Evolution ofSocialBehaviour,* Oxford, W. H. Freeman, 1996.
Hardin, Garrett, «The Tragedy of the Commons», *Science,* n° 162, 1968, págs. 1.243-1.248.
Harris, Judith, *The Nurture Assumption: Why Children Turn Out the Way They Do,* Nueva York, Touchstone (Simon and Schuster), 1998.
Hart, H. L. A. y A. M. Honoré, *Causation in the Law,* Oxford, Clarendon Press, 1959.
Haugeland, John, *Artificial Intelligence: The Very Idea,* Cambridge, MA, MIT Press, 1985.
—, *Having Thought: Essays in the Metaphysics of Mind,* Cambridge, MA, Harvard University Press, 1999.
Hofstadter, Douglas, *Le Ton Beau de Marot: In Praise of the Beauty of Language,* Nueva York, Basic Books, 1997.
Holmes, Bob, «Irresistible Illusions», *New Scientist,* vol. 159, n" 2.150, 1998, págs. 32-37.
Honderich, Ted, *A Theory of Determinism: The Mind, Neuroscience, and Life-Hopes,* Oxford, Oxford University Press, 1988.
Honoré, A. M., «Can and Can't», *Mind,* vol. 73, n° 292,1964, págs. 463-479.
Hooper, Lora V., Lynn Bry, Per G. Falk y Jeffrey I. Gordon, «Hostmicrobial Symbiosis in the Mammalian Intestine: Exploring an Internal Ecosystem», *BioEssays,* vol. 20, n° 4, 1998, págs. 336-343.
Hume, David (1739), *A Treatise of Human Nature,* edición a cargo de L. A. Selby-Bigge, Oxford, Clarendon Press, 1964 (trad, cast.: *Tratado de la naturaleza humana,* Madrid, Tecnos, 1988).
James, William (1897), *The Will to Believe and Other Essays,* Nueva York, Dover, reimpr. 1956 (trad, cast.: *La voluntad de creer: un debate sobre la ética de la creencia,* Madrid, Tecnos, 2003).

— (1907), *Pragmatism,* introducción de H. S. Thayer, Cambridge, MA, Harvard University Press, reimpr. 1975 (trad, cast.: *Pragmatismo: un nuevo nombre para viejas formas de pensar,* Madrid, Alianza, 2000).

Jensen, A. R., «g: Outmoded Theory or Unconquered Frontier?», *Creative Science and Technology,* n" 11, 1979, págs. 16-29.

Kane, Robert, *The Significance of Free Will,* Oxford, Oxford University Press, 1996.

—, «Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism »*Journal of Philosophy,* n° 96, 1999, págs. 217-240.

—, (comp.), *The Oxford Handbook of Free Will,* Nueva York, Oxford University Press, 2001.

Kant, Immanuel (1784), «Idea for a Universal History with a Cosmopolitan Purpose», en Immanuel Kant, *Kant's Political Writings,* edición a cargo de Hans Reiss, Cambridge, Cambridge University Press, reimpr. 1970.

Kass, Leon R, «Beyond Biology» (reseña de Brian Appleyard, *Staying Human in the Genetic Future), New York Times Book Review,* 23 de agosto de 1998, págs. 7-8.

Katz, Leonard D., «Toward Good and Evil: Evolutionary Approaches to Aspects of Human Morality», *Journal of Consciousness Studies,* vol. 7, n° 1-2, 2000; publicado también como Leonard D. Katz (comp.), *Evolutionary Origins of Morality: Cross-Disciplinary Perspectives,* Bowling Green, OH, Imprint Academic, 2000.

Kornhuber, H. H. y L. Deecke, «Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale», *Pflügers Arch. ges. Physiol.,* n°284, 1965, págs. 1-17.

Kripke, Saul, «Naming and Necessity», en D. Davidson y G. Harman (comps.), *Semantics of Natural Language,* Dordrecht, Reidel, 1972.

Laplace, Pierre-Simon (1814), *A Philosophical Essay on Probabilities,* Nueva York, Dover, reimpr. 1951 (trad, cast.: *Ensayo filosófico sobre las probabilidades,* Madrid, Alianza, 1985).

Leigh, E. G., *Adaptation and Diversity: Natural History and the Mathematics of Evolution,* San Francisco, Freeman, Cooper, 1971.

Lewis, David, *Counterfactuals,* Cambridge, MA, Harvard University Press, 1973.

—, «Causation as Influence», *journal of Philosophy,* n° 97,2000, págs. 182-197.

Lewontin, Richard, Steven Rose y Leon Kamin, *Not in Our Genes: Biology, Ideology, and Human Nature,* Nueva York, Pantheon, 1984 (trad, cast.: *No está en los genes,* Barcelona, Crítica, 1987).

Libet, Benjamin, «The Experimental Evidence for Subjective Referral of a Sensory Experience Backwards in Time: Reply to P. S. Churchland», *Philosophy of Science,* n° 48,1981, págs. 182-197.

—, «The Neural Time Factor in Conscious and Unconscious Mental Events», *Experimental and Theoretical Studies of Consciousness,* Simposio de la Fundación CIBA, n° 174, Chichester, Wiley, 1993.

—, «Neural Time Factors in Conscious and Unconscious Mental Function», en S. R. Hameroff, A. Kaszniak y A. Scott (comps.), *Toward a Science of Consciousness,* Cambridge, MA, MIT Press, 1996.

—, «Do We Have Free Will?», en Libet y otros, 1999, págs. 45-55.

Libet, Benjamin, Anthony Freeman y Keith Sutherland, *The Volitional Brain: Towards a Neuroscience of Free Will,* Thorverton, Imprint Academic, 1999.

Libet, Benjamin, C. A. Gleason, E. W Wright y D. K. Pearl, «Time of Conscious Intention to Act in Relation to Onset of Cerebral Activities (Readiness Potential); the Unconscious Initiation of a Freely Voluntary Act», *Brain,* n° 106, 1983, págs. 623-642.

MacKay, D. M., «On the Logical Indeterminacy of a Free Choice», *Mind,* n° 69, 1960, págs. 31-40.

MacKenzie, Robert Beverley, *The Darwinian Theory of the Transmutation of Species Examined* (publicado anónimamente «por un graduado de la Universidad de Cambridge»), Londres, Nisbet and Co., reproducido en *Athenaeum,* n° 2.102, 8 de febrero de 1868, pág. 217.

Mameli, Matteo, «Learning, Evolution, and the Icing on the Cake» (reseña de Avital y Jablonka, 2000), *Biology and Philosophy,* vol. 17, n° 1, 2002, págs. 141-153.

Marx, Karl (1867), *Capital, V* ed. inglesa, Moscú, Progress Publishers, 1887 (trad, cast.: *El capital,* Madrid, Alba, 1999).

Maxwell, Nicholas, *From Knowledge to Wisdom: A Revolution in the Aims and Methods of Science,* Oxford, Blackwell, 1984.

Maynard Smith, John (1982), «Models of Cultural and Genetic Change», en John Maynard Smith, *Games, Sex and Evolution,* Hemel Hempstead, Harvester, 1988.

Maynard Smith, John y Eörs Szathmáry, *The Major Transitions in Evolution,* Oxford, Freeman, 1995.

—, *The Origins of Life: From the Birth of Life to the Origin of Language,* Oxford, Oxford University Press, 1999 (trad, cast.: *Ocho hitos de la evolución: del origen de la vida a la aparición del lenguaje,* Barcelona, Tusquets, 2001).

McDonald, John F., «Transposable Elements, Gene Silencing and Macroevolution», *Trends in Ecology and Evolution,* n° 13, 1998, págs. 94-95.

McFarland, David, «Goals, No-Goals, and Own Goals», en Alan Montefiore y Denis Noble (comps.), *Goals, No-Goals, and Own Goals: A Debate on Goal-directed and Intentional Behaviour,* Londres, Unwin Hyman, 1989, págs. 39-57.

McGeer, Victoria, «Psycho-practice, Psycho-theory, and the Contrastive Case of Autism», *Journal of Consciousness Studies,* n° 8, 2001, págs. 109-132.

McGeer, Victoria y Philip Pettit, «The Self-Regulating Mind», *Language and Communication,* vol. 22, n° 3,2002, págs. 281-299.

McLaughlin, J. A., «Proximate Cause», *Harvard Law Review,* vol. 39, n° 149, 1925, pág. 155.

Mele, Alfred, *Autonomous Agents: From Self-Control to Autonomy*, Oxford, Oxford University Press, 1995.

Metcalfe, J. y W. Mischel, «A Hot/Cool System Analysis of Delay of Gratification: Dynamics of Willpower», *Psychological Review,* n° 106, 1999, págs. 3-19.

Milton, Katherine, «Civilization and Its Discontents», *Natural History,* marzo de 1992, págs. 37-42.

Moore, G. E., *Ethics,* Nueva York, H. Holt, 1912 (trad, cast.: *Etica,* Madrid, Encuentro, 2001).

Moore, Wayne R., *Foundations of Mechanical Accuracy,* Bridgeport, CT, Moore Special Tool Co., 1970.

Moya, Andrés y Enrique Font (comps.), *Evolution: From Molecules to Ecosystems,* Oxford, Oxford University Press, 1970.

Nesse, Randolph (comp.), *Evolution and the Capacity for Commitment,* Nueva York, Russell Sage, 2001.

Nozick, Robert, *Invariances: The Structure of the Objective World,* Cambridge, MA, Harvard University Press, 2001.

Pearl, Judea, *Causality: Models, Reasoning, and Inference,* Cambridge, Cambridge University Press, 2000.

Penrose, Roger, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics,* Oxford, Oxford University Press, 1989 (trad, cast.: *La nueva mente del emperador,* Barcelona, Grijalbo Mondadori, 1999).

—, *Shadows of the Mind: A Search for the Missing Science of Consciousness,* Nueva York, Oxford University Press, 1994 (trad, cast.: *Las sombras de la mente,* Barcelona, Crítica, 1996).

Pereboom, Derk, *Living without Free Will,* Cambridge, Cambridge University Press, 2001.

Pessin, Andrew y Sanford Goldberg (comps.), *The Twin Earth Chronicles,* Armonk, NY, M. E. Sharpe, 1996.

Pettit, Philip, *A Theory of Freedom: From the Psychology to the Politics of Agency,* Oxford, Oxford University Press, 2001.

Pinker, Steven, *How the Mind Works,* Nueva York, Norton, 1997 (trad, cast.: *Cómo funciona la mente,* Barcelona, Destino, 2001).

Popper, Karl, «Indeterminism in Quantum Physics and Classical Physics», *British journal for the Philosophy of Science,* n° 1, 1951, págs. 179-188.

Poundstone, William, *The Recursive Universe: Cosmic Complexity and the Limits of Scientific Knowledge,* Nueva York, Morrow, 1985.

Prentky, R. A., «Arousal Reduction in Sexual Offenders: A Review of Antiandrogen Interventions», *Sexual Abuse: A Journal of Research and Treatment,* n° 9, 1997, págs. 335-348.

Pynchon, Thomas, *Gravity's Rainbow,* Nueva York, Viking, 1973 (trad, cast.: *El arco iris de la gravedad,* Barcelona, Tusquets, 2002).

Quine, W. V. O., «Propositional Objects», en W. V. O. Quine, *Ontological Relativity and Other Essays,* Nueva York, Columbia University Press, 1969, pags. 147-155 (trad, cast.: *Relatividad ontoldgica y otros ensayos,* Madrid, Tecnos, 1974).

Quinsey, Vernon L., Grant T. Harris, Mamie E. Rice y Catherine A. Cormier, *Violent Offenders: Appraising and Managing Risk,* Washington, D. C., American Psychological Association, 1998.

Raffman, Diana, «Vagueness and Context Relativity», *Philosophical Studies,* vol. 81, n° 2-3,1996, pags. 175-192.

Raine, Adrian y otros, «Birth Complications Combined with Early Maternal Rejection at Age 1 Year Predispose to Violent Crime at Age 18 Years», *Archives of General Psychiatry,* n° 51, 1994, pags. 984-988.

Ramachandran, Vilayanur, reproducido en *New Scientist,* 5 de septiembre de 1998, pag. 35.

Rawls, John, *A Theory of Justice,* Cambridge, MA, Harvard University Press, 1971 (trad, cast.: *Teoria de la justicia,* Madrid, Fondo de Cultura Economica, 1997).

Ridley, Mark, *Animal Behaviour, T* ed., Boston, Blackwell Scientific Publications, 1995.

Ridley, Matt, *The Origins of Virtue,* Nueva York, Viking, 1996.

—, *Genome: The Autobiography of a Species in 23 Chapters,* Londres, Fourth Estate, 1999.

Rosier, A. y E. Witztum, «Treatment of Men with Paraphilia with a Long-acting Analogue of Gonadotropin-releasing Hormone», *New England Journal of Medicine,* n° 338,1998, pags. 416-422.

Ross, Don y Paul Dumouchel, «Emotions as Strategic Signals», accesible en http://www.commerce.uct.ac.za/economics/staff/personalpages/dross/emotelO.rtf

Ryle, Gilbert, *The Concept of Mind,* Londres, Hutchinson, 1949.

Sanford, David, «Infinity and Vagueness», *Philosophical Review,* n° 84, 1975, pags. 520-535.

Sartre, Jean Paul (1943), *Being and Nothingness,* Nueva York, Simon and Schuster, col. «Philosophical Library», reimpr. 1966 (trad, cast.: *El ser y la nada,* Madrid, Alianza, 1989).

Sellars, Wilfrid, «Empiricism and the Philosophy of Mind», en Wilfrid Sellars, *Science, Perception, and Reality,* Londres, Routledge and Kegan Paul, 1963, pags. 127-196.

Sen, Amartya, *Development as Freedom,* Nueva York, Knopf, 1999 (trad, cast.: *Desarrollo y libertad,* Barcelona, Planeta, 2000).

Skyrms, Brian, «Sex and Justice», *Journal of Philosophy,* n° 91, 1994a, pags. 305-320.

—, «Darwin Meets *The Logic of Decision*-. Correlation in Evolutionary Game Theory», *Philosophy of Science,* n° 62,1994b, pags. 503-528.

—, *Evolution of the Social Contract,* Nueva York, Cambridge University Press, 1996.

Slote, Michael, «Ethics without Free Will», *Social Theory and Practice,* n° 16, 1990, págs. 369-383.

Sober, Elliott y David Sloan Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behavior,* Cambridge, MA, Harvard University Press, 1998.

Sperber, Dan (comp.), *The Epidemiology of Ideas,* numero especial de *The Monist, vol.* 84, n° 3,2001.

Sterelny, Kim y Paul E. Griffiths, *Sex and Death: An Introduction to Philosophy of Biology,* Chicago, University of Chicago Press, 1999.

Stich, Stephen, *Deconstructing the Mind,* Oxford, Oxford University Press, 1996.

Suber, Peter, «The Paradox of Liberation», 1992, accesible en http://www.earlham.edu/~peters/writing/liber.htm

Taylor, Christopher y Daniel Dennett, «Who's Afraid of Determinism? Rethinking Causes and Possibilities», en Robert Kane (comp.), *Oxford Handbook of Free Will,* Nueva York, Oxford University Press, 2001.

Thompson, Adrian, P. Layzell y R. S. Zebulum, «Explorations in Design Space: Unconventional Electronics Design through Artificial Evolution», *IEEE (Institute of Electrical and Electronics Engineers) Transactions on Evolutionary Computation,* n° 3,1999, págs. 167-196.

Turing, Alan, «On Computable Numbers, with an Application to the *Entscheidungsproblem», Proceedings of the London Mathematical Society*, vol. 2, n° 42, 1936, págs. 230-265.

Van Inwagen, Peter, *An Essay on Free Will,* Oxford, Clarendon Press, 1983.

Velleman, David, «What Happens When Someone Acts?», *Mind,* n° 101, 1992, págs. 461-481.

Wagensberg, Jorge, «Complexity versus Uncertainty: The Question of Staying Alive», *Biology and Philosophy,* n° 15, 2000, págs. 493-508.

Weber, Bruce y David Depew (comps.), *Evolution and Learning: The Baldwin Effect Reconsidered,* Cambridge, MA, MIT Press, 2002.

Wegner, Daniel, *The Illusion of Conscious Will,* Cambridge, MA, MIT Press, 2002.

White, Stephen L,, *The Unity of the Self,* Cambridge, MA, MIT Press, 1991.

Whitehead, Alfred North (1933), *Adventures of Ideas,* Nueva York, Macmillan, reimpr. 1967.

Wiggins, David, «Natural and Artificial Virtues: A Vindication of Hume's Scheme», en Roger Crisp (comp.), *How Should One Live? Essays on the Virtues,* Oxford, Clarendon Press, 1996, págs. 131-140.

Williams, George, «Reply to Comments on "Huxley's Evolution and Ethics in Sociobiological Perspective"», *Zygon,* vol. 23, n° 4,1988, págs. 437-438.

Wolfe, Jeremy M., George A. Alvarez y Todd S. Horowitz, «Attention Is Fast but Volition Is Slow», *Nature,* n° 406,2000, pág. 691.

Wolfe, Tom, *Hooking Up,* Nueva York, Farrar, Straus and Giroux, 2000.

Wright, Robert, *The Moral Animai: The New Science of Evolutionary Psychology,* Nueva York, Pantheon, 1994.
—, *Nonzero: The Logic ofHunian Destiny,* Nueva York, Pantheon, 2000.
Zahavi, Amotz, «The Theory of Signal Selection and Some of Its Implications», en V. P. Delfino (comp.), *International Symposium on Biological Evolution, Bari, 9-14 April 1985,* Bari, Adriatici Editrici, 1987, pàgs. 305-327.

Indice analítico y de nombres