Bombas de intuición y otras herramientas de pensamiento

Daniel C. Dennett

FILOSOFÍA

DANIEL C. DENNETT (Boston, 1942) es un destacado filósofo estadunidense cuyas líneas de investigación se centran en la filosofía de la mente y la filosofía de la ciencia, en particular la biología evolutiva y las ciencias cognitivas. Actualmente es profesor y codirector del Centro de Estudios Cognitivos en la Tufts University.

•

Un filósofo con una originalidad, rigor e ingenio excepcionales.

JIM HOLT
The Wall Street Journal

El mejor de nuestros filósofos actuales, el siguiente Bertrand Russell. A diferencia de los filósofos tradicionales, Daniel es un estudioso de las neurociencias, la lingüística, la inteligencia artificial, la ciencia computacional y la psicología. Está redefiniendo y trasformando el papel de los filósofos.

MARVIN MINSKY

Uno de los pensadores más originales de nuestro tiempo.

MICHAEL SHERMER
Science

La más aguda, ingeniosa y sofisticada demostración de la forma en que los problemas derivados de la conciencia humana están relacionados hoy en día con la teoría de la evolución.

Rara vez encontramos un filósofo analítico que haga justicia a su subdisciplina al combinar la precisión profesional con un espíritu romántico, una imaginación vívida y sentido del humor. Definitivamente es uno de los filósofos más originales y agradables de leer.

RICHARD RORTY

Bombas de intuición y otras herramientas de pensamiento

Traducción
LAURA LECUONA

Daniel C. Dennett

Bombas de intuición y otras herramientas de pensamiento

Sección de Obras de Filosofía

Primera edición en inglés, 2013
Primera edición en español, 2015
Primera edición electrónica, 2015

Diseño de portada: Teresa Guzmán Romero

Título original: *Intuition Pumps and Other Tools for Thinking*
© 2013 by Daniel C. Dennett . Todos los derechos reservados.

D. R. © 2015, Fondo de Cultura Económica
Carretera Picacho-Ajusco, 227; 14738 México, D. F.
Empresa certificada ISO 9001:2008

www.fondodeculturaeconomica.com

Comentarios:
editorial@fondodeculturaeconomica.com
Tel. (55) 5227-4672

Se prohíbe la reproducción total o parcial de esta obra, sea cual fuere el medio. Todos los contenidos que se incluyen tales como características tipográficas y de diagramación, textos, gráficos, logotipos, iconos, imágenes, etc. son propiedad exclusiva del Fondo de Cultura Económica y están protegidos por las leyes mexicana e internacionales del copyright o derecho de autor.

ISBN 978-607-16-3153-4 (mobi)

Hecho en México - *Made in Mexico*

Para la Universidad Tufts, mi hogar académico

Sumario

Prefacio

I. Introducción. ¿Qué es una bomba de intuición?
II. Doce herramientas de pensamiento generales
III. Herramientas para pensar acerca del significado o contenido
IV. Intermedio sobre computadoras
V. Más herramientas sobre el significado
VI. Herramientas para pensar acerca de la evolución
VII. Herramientas para pensar acerca de la conciencia
VIII. Herramientas para pensar acerca del libre albedrío
IX. ¿Cómo es ser un filósofo?
X. Usa las herramientas. Esfuérzate más
XI. Lo que quedó fuera

Apéndice. Soluciones a los problemas de la máquina de registro
Fuentes documentales
Bibliografía
Índice analítico
Índice general

Prefacio

La Universidad Tufts ha sido mi hogar académico durante más de 40 años, y para mí siempre ha sido perfecta, como la sopa de Ricitos de Oro: no demasiado agobiante, no demasiado consentidora, con colegas brillantes de los que se puede aprender, un mínimo de *prima donnas* académicos, buenos estudiantes lo suficientemente serios para merecer atención sin pretender tener derecho a mantenimiento las 24 horas, una torre de marfil profundamente comprometida con la solución de problemas en el mundo real. Desde la creación del Centro de Estudios Cognitivos, en 1986, Tufts ha apoyado mi investigación, me ha ahorrado buena parte de los sufrimientos y obligaciones asociados a las becas y me ha dado una notable libertad para trabajar con gente de otros campos, ya sea que viaje lejos a talleres, laboratorios y conferencias, o traiga académicos visitantes y otros al centro. Este libro muestra en qué he estado metido todos estos años.

En la primavera de 2012 hice vuelos de prueba con un primer borrador de los capítulos en un seminario que impartí en el Departamento de Filosofía de Tufts. Ésa ha sido mi costumbre por años, pero en esta ocasión quería que los estudiantes me ayudaran a hacer el libro tan accesible al no iniciado como fuera posible, así que excluí a los estudiantes de posgrado y a los alumnos de filosofía, y limité la clase a solamente una docena de intrépidos estudiantes de primer año: los primeros 12 voluntarios —de hecho 13, debido a una torpeza administrativa—. Nos guiamos mutuamente en un divertido viaje a través de los temas, mientras ellos aprendían que realmente podían plantar cara al profesor y yo aprendía que realmente podía remontarme más atrás y explicar todo mejor. Agradezco, pues, a mis jóvenes colaboradores por su valentía, imaginación, energía y entusiasmo: Tom Addison, Nick Boswell, Tony Cannistra, Brendan Fleig Goldstein, Claire Hirschberg, Caleb Malchik, Carter Palmer, Amar Patel, Kumar Ramanathan, Ariel Rascoe, Nikolai Renedo, Mikko Silliman y Eric Tondreau.

Luego di a leer el segundo borrador que surgió de ese seminario a mis queridos amigos Bo Dahlbom, Sue Stafford y Dale Peterson, quienes me ofrecieron todavía más comentarios sinceros y sugerencias, la mayoría de los cuales he atendido, y a mi editor, Drake McFeely, y a su apto asistente, Brendan Curry, en W. W. Norton, quienes también son autores de muchas mejoras, por las que les estoy agradecido. Gracias especialmente a Teresa Salvato, coordinadora de programas del Centro de Estudios Cognitivos, quien de innumerables maneras contribuyó directamente al proyecto en su totalidad y ayudó indirectamente al administrar el centro y mis viajes de manera tan eficiente que pude

dedicar más tiempo y energía a construir y emplear mis herramientas de pensamiento.

Finalmente, como siempre, amor y agradecimiento a mi esposa, Susan. Durante 50 años hemos sido un equipo, y ella es tan responsable como yo de lo que juntos hemos hecho.

DANIEL C. DENNETT
Blue Hill, Maine, agosto de 2012

I. Introducción

¿Qué es una bomba de intuición?

> No puedes hacer mucho trabajo de carpintería con las puras manos, y no puedes hacer mucho trabajo de pensamiento con el puro cerebro.
>
> BO DAHLBOM

Pensar es difícil. Pensar sobre algunos problemas es tan difícil que tan sólo pensar en pensar en ellos te puede provocar un dolor de cabeza. Mi colega el neuropsicólogo Marcel Kinsbourne sugiere que siempre que pensar nos parece difícil es porque el pedregoso camino a la verdad compite con seductores caminos más sencillos que resultan ser callejones sin salida. La mayor parte del esfuerzo que implica pensar consiste en resistir estas tentaciones. Ellas siguen abordándonos y tenemos que mantenernos firmes en nuestra tarea. *Puf*.

Hay una famosa anécdota sobre John von Neumann, el matemático y físico que convirtió la idea de Alan Turing (lo que ahora llamamos máquina de Turing) en una computadora electrónica real (lo que ahora llamamos máquina de Von Neumann, como tu computadora portátil o tu teléfono inteligente). Von Neumann era un pensador virtuoso, legendario por su capacidad como de rayo para hacer cálculos prodigiosos en la cabeza. Dice una anécdota —y, como todas las anécdotas famosas, ésta tiene muchas versiones— que un día lo abordó un colega con un acertijo que tenía dos caminos a una solución: un cálculo laborioso y complicado, y una solución elegante de esas que hacen decir «¡Ajá!». Este colega tenía una teoría: en casos así, los matemáticos encuentran la solución laboriosa, mientras que los físicos (más flojos, pero más listos) se detienen un momento y encuentran la solución rápida y fácil. ¿Qué solución encontraría Von Neumann? Te será familiar esta clase de acertijo: dos trenes, separados por 160 kilómetros, se están acercando sobre la misma vía: uno va a 50 kilómetros por hora, el otro va a 30 kilómetros por hora. Un pájaro que vuela a 190 kilómetros por hora parte del tren A (cuando están a 160 kilómetros de distancia), vuela al tren B, al llegar se da media vuelta y vuela de regreso al tren A que se aproxima, y así sucesivamente, hasta que los dos trenes chocan. ¿Cuánto ha volado el pájaro cuando ocurre la colisión? «Trescientos ochenta kilómetros», respondió Von Neumann casi al instante. «Demonios —dijo su colega—, predije que lo harías del modo difícil, sumando las series infinitas.» «Oh —exclamó Von Neumann avergonzado, golpeándose la frente—, ¡hay una manera fácil!» (Pista: ¿cuánto tiempo pasa hasta que los trenes chocan?)

Algunas personas, como Von Neumann, son genios naturales que pasan sin dificultades a través de las marañas más enredadas; otros necesitan aplicarse más pero

están bendecidos con una provisión heroica de «fuerza de voluntad» que les ayuda a mantener el rumbo en su obstinada búsqueda de la verdad. Y luego estamos los demás, no prodigios del cálculo y un poco flojos, pero también aspiramos a comprender aquello a lo que nos enfrentamos. ¿Qué podemos hacer? Podemos usar herramientas de pensamiento por docenas. Estas accesibles prótesis que extienden la imaginación y mantienen el foco nos permiten pensar de manera confiable e incluso elegante sobre preguntas realmente difíciles. Este libro es una colección de mis herramientas de pensamiento favoritas. No sólo las describiré; pretendo *usarlas* para llevar tu mente con suavidad por todo el camino a través de un territorio incómodo hasta llegar a una visión bastante radical del significado, la mente y el libre albedrío. Comenzaremos con algunas herramientas simples y generales que tienen aplicación en toda clase de temas. Algunas de ellas son conocidas, pero de otras no se habla mucho. Luego te presentaré algunas herramientas para fines ciertamente muy especiales, concebidas para hacer estallar una u otra idea seductora específica y despejar así el camino para salir de un surco profundo que aún atrapa y desconcierta a los expertos. También abordaremos y desarmaremos una variedad de malas herramientas de pensamiento, recursos persuasivos mal concebidos que si no tienes cuidado te pueden descarriar. Llegues o no cómodamente al destino que propongo —y ya sea que te quedes ahí conmigo o no—, el viaje te equipará con nuevas maneras de pensar acerca de los temas abordados y nuevas maneras de pensar acerca del pensamiento.

El físico Richard Feynman fue un genio tal vez aún más legendario que Von Neumann, y sin duda estaba dotado de un cerebro de primera categoría; pero también le encantaba divertirse, y podemos estar agradecidos de que disfrutara particularmente al revelar los secretos del oficio que usaba para hacerse la vida más fácil. No importa qué tan listo seas, eres más listo si tomas los caminos fáciles cuando están disponibles. Sus libros autobiográficos, *¿Está usted de broma, señor Feynman?* y *What Do You Care What Other People Think?,* deberían estar en la lista de lecturas obligatorias de cualquier aspirante a pensador, pues tienen muchas pistas sobre cómo domar los problemas más peliagudos, e incluso sobre cómo deslumbrar a un público con farsas si nada mejor viene a la mente. Inspirado por la riqueza de observaciones útiles en sus libros y por su franqueza al revelar cómo trabajaba su mente, decidí intentar con mis propias manos un proyecto similar, menos autobiográfico y con la ambiciosa meta de convencer al lector para que piense sobre estos temas *a mi manera.* Me tomaré muchas molestias para engatusarlo y hacerlo salir de algunas de sus convicciones más firmes, pero sin nada bajo la manga. Uno de mis propósitos principales es revelar sobre la marcha exactamente qué estoy haciendo y por qué.

Como todos los artesanos, un herrero necesita herramientas, pero —según una vieja y de hecho casi extinta observación— los herreros tienen la singularidad de que hacen

sus propias herramientas. Los carpinteros no hacen sus sierras y martillos, los sastres no hacen sus tijeras y agujas, y los plomeros no hacen sus llaves, pero los herreros pueden hacer sus martillos, pinzas, yunques y cinceles con su materia prima, el hierro. ¿Y qué hay de las herramientas para pensar? ¿Quién las hace? ¿Y de qué están hechas? Los filósofos han hecho algunas de las mejores, y a partir de nada más que ideas, estructuras útiles de información. René Descartes nos dio las *coordenadas cartesianas,* los ejes *x* y *y,* sin las cuales el *cálculo* —herramienta de pensamiento por antonomasia inventada simultáneamente por Isaac Newton y el filósofo Gottfried Wilhelm Leibniz— sería casi impensable. Blaise Pascal nos dio la *teoría de la probabilidad* para que fácilmente podamos calcular las probabilidades de varias apuestas. El reverendo Thomas Bayes era también un talentoso matemático, y nos dio el *teorema de Bayes,* la columna vertebral del pensamiento estadístico bayesiano. Sin embargo, la mayoría de las herramientas que aparecen en este libro son más simples, no las máquinas precisas y sistemáticas de la ciencia y las matemáticas, sino las herramientas manuales de la mente. Entre ellas están las siguientes:

- *Etiquetas.* A veces tan sólo crear un nombre vivaz para algo te ayuda a darle seguimiento mientras le das vueltas en la mente tratando de entenderlo. Como veremos, entre las etiquetas más útiles están las de advertencia, o alarmas, que nos alertan sobre fuentes probables de error.
- *Ejemplos.* Algunos filósofos piensan que usar ejemplos en su obra, si bien no es totalmente tramposo, al menos está fuera de lugar, tal como los novelistas rehúsan que sus novelas tengan ilustraciones. Los novelistas se enorgullecen de hacerlo todo con palabras, y los filósofos se enorgullecen de hacerlo todo con generalizaciones abstractas cuidadosamente confeccionadas presentadas en un orden riguroso, tan cercano a las demostraciones matemáticas como puedan lograrlo. Bien por ellos, pero no pueden esperar que yo recomiende su obra más que a unos cuantos estudiantes destacados. Es, simplemente, más difícil de lo que tendría que ser.
- *Analogías y metáforas.* Representar las características de una cosa compleja aludiendo a las características de otra cosa compleja que (crees que) ya conoces es una conocida y poderosa herramienta de pensamiento, pero es tan poderosa que a menudo descarría a los pensadores por el mal camino cuando sus imaginaciones caen presas de una analogía traicionera.
- *Andamiajes.* Puedes cubrir de pizarra un tejado, pintar una casa o arreglar una chimenea con la sola ayuda de una escalera, moviéndola y ascendiendo, moviéndola y ascendiendo de nuevo, teniendo acceso a solamente una pequeña parte del trabajo cada vez, pero a menudo resulta mucho más fácil que al principio te tomes el tiempo de armar un andamiaje resistente que te permita moverte con rapidez y seguridad alrededor de todo el proyecto. Varias de las herramientas de pensamiento más valiosas de este libro son ejemplos de

andamiaje, que toma cierto tiempo levantar pero luego permiten que una variedad de problemas se encaren conjuntamente, y sin andar moviendo escaleras de un lado a otro.
- Y, finalmente, la clase de experimentos mentales que he llamado *bombas de intuición.*

No sorprende que los experimentos mentales estén entre las herramientas favoritas de los filósofos. ¿Quién necesita un laboratorio cuando puede encontrar la respuesta a su pregunta mediante alguna deducción ingeniosa? Los científicos, de Galileo a Einstein y más allá, también han usado experimentos mentales con buenos resultados, así que no son herramientas nada más de los filósofos. Algunos experimentos mentales pueden analizarse como argumentos rigurosos, a menudo de la forma *reductio ad absurdum,*[1] consistente en tomar las premisas del oponente y derivar una contradicción formal (un resultado absurdo), con lo que se muestra que no pueden ser todas ellas correctas. Una de mis favoritas es la demostración atribuida a Galileo de que las cosas pesadas no caen más rápido que las cosas ligeras (cuando la fricción es insignificante). Si así fuera, sostenía, entonces como la piedra pesada *A* caería más rápido que la piedra ligera *B,* si atáramos *B* a *A,* la piedra *B* actuaría como una rastra, desacelerando a *A;* pero *A* atada a *B* es más pesada que *A* sola, así que las dos juntas deberían también caer más rápido que *A* por sí sola. Hemos concluido que atar *B* a *A* daría lugar a algo que caería más rápido y más lento que *A* por sí sola, lo cual es una contradicción.

Otros experimentos mentales son menos rigurosos, pero a menudo igual de efectivos: pequeñas historias ideadas para provocar una intuición sincera que haga golpear la mesa («¡Sí, por supuesto, tiene que ser así!»), sobre cualquier tesis que se esté defendiendo. Las he llamado *bombas de intuición.* Acuñé el término en la primera de mis críticas públicas al famoso experimento mental de John Searle de la habitación china,[2] y algunos pensadores concluyeron que yo pretendía que el término fuera desdeñoso o despectivo. ¡Por el contrario, me encantan las bombas de intuición! Esto es, algunas bombas de intuición son excelentes, otras son ambiguas, y sólo unas cuantas son categóricamente engañosas. Las bombas de intuición han sido una fuerza dominante en la filosofía durante siglos. Son la versión filosófica de las fábulas de Esopo, que se han reconocido como maravillosas herramientas de pensamiento desde antes de que hubiera filósofos.[3] Si alguna vez cursaste filosofía en la universidad, probablemente estuviste expuesto a algunas clásicas, como la caverna de Platón, en *La república,* en la que la gente está encadenada y solamente puede ver las sombras de las cosas reales que se proyectan sobre el muro de la caverna; o a su ejemplo, en el *Menón,* de enseñar geometría al niño esclavo. Y luego tenemos el genio maligno de Descartes, que engaña a Descartes al hacerle creer en un mundo completamente ilusorio —el experimento mental de la

realidad virtual originario— y el estado de naturaleza de Hobbes, en el que la vida es desagradable, salvaje y breve. No tan famosas como «El pastor mentiroso» o «La cigarra y la hormiga», pero aun así ampliamente conocidas, cada una está ideada para bombear algunas intuiciones. La caverna de Platón pretende iluminarnos acerca de la naturaleza de la percepción y la realidad, y el niño esclavo está pensado para ilustrar nuestro conocimiento innato; el genio maligno es el máximo generador de escepticismo, y la parábola de Hobbes apunta a nuestra mejora con respecto al estado de naturaleza cuando pactamos un contrato para formar una sociedad. Éstas son las melodías duraderas de la filosofía, con la perdurabilidad que asegura que los estudiantes las recuerden, muy vívidamente y con exactitud, años después de haber olvidado los intrincados argumentos y análisis circundantes. Una buena bomba de intuición es más sólida que cualquier versión de ella. Examinaremos una variedad de bombas de intuición contemporáneas, algunas de ellas defectuosas, y la meta será entender para qué son buenas, cómo funcionan, cómo usarlas e incluso cómo hacerlas.

He aquí un breve ejemplo sencillo: el carcelero caprichoso. Cada noche espera a que todos los presos estén completamente dormidos y entonces hace una ronda para quitar los cerrojos de todas las puertas y las deja abiertas horas y horas. Pregunta: ¿son libres los presos? ¿Tienen una *oportunidad* de irse? En realidad, no. ¿Por qué no? Otro ejemplo: las joyas en el bote de basura. Resulta que hay una fortuna en joyería tirada en el bote de basura de la banqueta por la que vas caminando una noche. Parecería que tienes una oportunidad de oro de volverte rico, excepto que no es precisamente de oro porque es una oportunidad *escasa,* una que sería extremadamente poco probable que reconocieras y por ende que actuaras en función de ella, o que siquiera tomaras en cuenta. Estos dos simples escenarios bombean intuiciones que de otra forma podrían no ser obvias: la importancia de obtener información puntual acerca de las oportunidades genuinas, con la antelación suficiente para que la información cause que la tomemos en cuenta a tiempo para poder hacer algo al respecto. En nuestro ímpetu por hacer elecciones «libres», no causadas por «fuerzas externas» —o así nos gusta creerlo—, tendemos a olvidar que no quisiéramos estar completamente apartados de tales fuerzas; el libre albedrío no *aborrece* que estemos incorporados a un rico contexto causal; de hecho, *requiere* que así sea.

Espero que te quedes con la impresión de que hay más que decir sobre ese tema. Estas pequeñas bombas de intuición plantean vívidamente un problema, pero no resuelven nada… aún (más adelante se dedica una sección completa al libre albedrío). Tenemos que adquirir práctica en el arte de tratar cautelosamente esas herramientas, mirando por dónde pisamos y atentos a los obstáculos. Si pensamos en una bomba de intuición como una herramienta de convencimiento cuidadosamente diseñada, veremos que tal vez nos recompense hacer la *ingeniería inversa* de la herramienta, para revisar

todas sus partes móviles y conocer sus funciones.

Cuando Doug Hofstadter y yo escribimos *The Mind's I,* allá por 1982, se le ocurrió el consejo perfecto en relación con esto: piensa en la bomba de intuición como si fuera una herramienta con muchos ajustes y «gira todas las perillas» para ver si al considerar algunas variaciones siguen bombeándose las mismas intuiciones.

Así, pues, identifiquemos, y giremos, las perillas del carcelero caprichoso. Asumamos —mientras no se demuestre lo contrario— que cada parte tiene una función, y veamos qué función es ésa remplazándola con otra parte, o transformándola ligeramente.

1. Cada noche
2. espera
3. a que todos los presos
4. estén completamente dormidos
5. y entonces hace una ronda para quitar los cerrojos
6. de todas las puertas
7. y las deja abiertas horas y horas.

He aquí una de muchas variaciones que podríamos considerar:

Una noche les ordenó a sus guardias que drogaran a uno de los presos y, después de hacerlo, accidentalmente dejaron la puerta de la celda de ese preso sin cerrojo durante una hora.

Cambia mucho el sabor del escenario, ¿no es así? ¿Cómo? Sigue haciendo su señalamiento central (¿no es así), pero de manera menos efectiva. La gran diferencia parece residir entre estar naturalmente dormido —*podrías* despertarte en cualquier minuto— y estar drogado o perdido. Otra diferencia —«accidentalmente»— subraya el papel de la intención o inadvertencia por el carcelero o los guardias. La repetición («cada noche») parece cambiar las probabilidades a favor de los presos. ¿Por qué y cómo importan las probabilidades? ¿Cuánto pagarías por *no* tener que participar en una lotería en la que millones de personas tuvieran boletos y al «ganador» le pegaran un tiro? ¿Cuánto pagarías por *no* tener que jugar a la ruleta rusa con un revólver de seis tiros? (Aquí estamos usando una bomba de intuición para iluminar otra: un truco para recordar.)

Otras perillas son menos obvias. En secreto, el Anfitrión Diabólico echa llave a la puerta de las habitaciones de sus huéspedes mientras duermen. La Gerente del Hospital, preocupada por la posibilidad de un incendio, de noche mantiene sin llave las puertas de todos los cuartos y salas, pero no lo informa a los pacientes, pues piensa que dormirán

mejor si no lo saben. ¿O qué tal si la cárcel es un poco más grande de lo habitual, digamos, del tamaño de Australia? No puedes quitar o echar cerrojos en todas las puertas de Australia. ¿Qué es lo que cambia aquí?

Esta cautela consciente con la que deberíamos abordar cualquier bomba de intuición es de suyo una importante herramienta para pensar, la táctica favorita del filósofo: «ponerse *meta*-», pensar acerca del pensamiento, hablar acerca del habla, razonar acerca del razonamiento. El *meta*lenguaje es el lenguaje que usamos para hablar acerca de otro lenguaje, y la *meta*ética es un examen a vista de pájaro de las teorías éticas. Como una vez le dije a Doug: «Cualquier cosa que puedas hacer, yo puedo hacerla *meta*-». Todo este libro es, desde luego, un ejemplo de ponerse *meta-,* pues explora cómo pensar cuidadosamente acerca de los métodos para pensar cuidadosamente (acerca de los métodos para pensar cuidadosamente, etc.).[4] Él recientemente presentó una lista de sus pequeñas herramientas de mano favoritas:[5]

– persecuciones inútiles *[wild goose chases]*
– pegajosidad *[tackiness]*
– artimañas *[dirty tricks]*
– uvas verdes *[sour grapes]*
– arremangarse *[elbow grease]*
– pies de barro *[feet of clay]*
– tipos impredecibles *[loose cannons]*
– chiflados *[crackpots]*
– de dientes para fuera *[lip service]*
– pan comido *[slam dunks]*
– retroalimentación *[feedback]*

Si estas expresiones te resultan familiares, no son para ti «simples palabras»; cada una es una herramienta cognitiva abstracta, del mismo modo que *división larga* o *sacar el promedio* son herramientas; cada una desempeña un papel en una amplia gama de contextos, con lo que se hace más fácil formular hipótesis para probar, con lo que se hace más fácil reconocer patrones inadvertidos en el mundo, que ayudan al usuario a buscar semejanzas importantes, y así sucesivamente. Cada palabra de tu vocabulario es una herramienta de pensamiento simple, pero algunas son más útiles que otras. Si alguna de estas expresiones no forma parte de tu juego de herramientas, tal vez quieras adquirirlas. Cuando estés equipado con ellas podrás pensar pensamientos que de otro modo sería relativamente difícil formular. Desde luego, como dice el dicho, cuando tu única herramienta es el martillo todo parece clavo, y puede abusarse de cada una de estas herramientas.

Veamos una sola de ellas: las uvas verdes. Viene de la fábula de Esopo «La zorra y las uvas» y llama la atención sobre cómo en ocasiones la gente finge no interesarse por algo que no puede obtener, menospreciándolo. Fíjate cuánto puedes expresar sobre lo que alguien acaba de decir simplemente preguntando «¿Uvas verdes?» Con eso puedes hacer que considere una posibilidad que de otro modo tal vez le habría pasado inadvertida, y eso podría inspirarlo eficazmente a revisar su pensamiento o a reflexionar sobre el tema desde una perspectiva más amplia… o podría insultarlo muy eficazmente (las herramientas también pueden usarse como armas). La moraleja de la historia es tan conocida que es posible que hayas olvidado el cuento que llevó a ella, o que hayas perdido contacto con sus sutilezas, si es que importan (y a veces no importan).

Adquirir herramientas y utilizarlas sabiamente son habilidades bien diferenciadas, pero tienes que empezar por adquirir las herramientas o fabricarlas tú mismo. Muchas de las herramientas de pensamiento que presento aquí son de mi propia invención, pero otras las he adquirido de otras personas, y reconoceré a sus inventores a su debido tiempo.[6] Ninguna de las herramientas de la lista de Doug es de su invención, pero él ha aportado algunos magníficos especímenes a mi juego, como el **pensar fuera de la caja** y el **sphexismo.**

Algunas de las herramientas de pensamiento más poderosas son matemáticas, pero, más allá de mencionarlas, no les dedicaré mucho espacio, pues éste es un libro que celebra el poder de las herramientas *no* matemáticas, las herramientas *informales,* las de la prosa y la poesía, si tú quieres, un poder que los científicos a menudo subestiman. Queda claro por qué. En primer lugar, en las revistas de investigación hay una cultura de escritura científica que favorece presentaciones impersonales y austeras de los temas con un mínimo de floritura, retórica y alusión (y de hecho insiste en que así sea). Hay una buena razón para la implacable monotonía en las páginas de nuestras revistas científicas más serias. Como uno de mis sinodales del doctorado, el neuroanatomista J. Z. Young, me escribió en 1965 al objetar la prosa un tanto extravagante de mi tesis de Oxford (en filosofía, no neuroanatomía), el inglés se estaba convirtiendo en la lengua internacional de la ciencia, y es deber de los hablantes nativos del inglés escribir obras que pueda leer un «chino pasciente *[sic]* con ayuda de un buen diccionario». Los resultados de esta disciplina autoimpuesta hablan por sí mismos: ya seas un científico chino, alemán, brasileño o incluso francés, insistes en publicar tu obra más importante en inglés, en el inglés más limitado, que pueda traducirse con la menor dificultad, y echas mano lo menos posible de alusiones culturales, matices, juegos de palabras o incluso metáforas. El nivel de comprensión mutua que ha alcanzado este sistema internacional es inestimable, pero tiene un precio: una parte del pensamiento que debe llevarse a cabo aparentemente *requiere* que se asalten las metáforas, que se pellizque a la imaginación, que se ataquen las barricadas de las mentes cerradas con todo tipo de artificios, y si una parte no puede

traducirse con facilidad, entonces simplemente tendré que esperar, por un lado, a que haya traductores virtuosos, y por otro, a que los científicos del mundo hablen el inglés cada vez con más soltura.

Otra razón por la que a menudo los científicos desconfían de las discusiones teóricas que se mantienen con «puras palabras» es que reconocen que hacer la crítica de un argumento que no está formulado en ecuaciones matemáticas es mucho más difícil, y casi siempre menos concluyente. Se puede confiar en el lenguaje de las matemáticas para dar una impresión de contundencia. Es como la red de un aro de baloncesto: elimina motivos de desacuerdo y ayuda a determinar si la pelota pasó por ahí o no (cualquiera que haya jugado con un aro sin red en la cancha de un parque sabrá lo difícil que es distinguir entre un tiro fallido y una canasta). Sin embargo, a veces los temas son demasiado escurridizos y enigmáticos para que las matemáticas puedan domarlos.

Siempre me ha parecido que si no puedo explicarle lo que estoy haciendo a un grupo de estudiantes de licenciatura inteligentes, en realidad yo mismo no lo entiendo. Este reto ha dado forma a todos mis escritos. Algunos profesores de filosofía anhelan impartir seminarios avanzados solamente a estudiantes de posgrado. Yo no. A menudo los estudiantes de posgrado están demasiado ansiosos por demostrarles a los demás y a sí mismos que son unos expertos: entonces blanden hábilmente y con aire de suficiencia la jerga de su especialidad, con lo que confunden a los forasteros (así es como se convencen a sí mismos de que para lo que están haciendo se requiere pericia), y lucen su habilidad para transitar por los argumentos técnicos más tortuosos (y torturantes) sin perderse. La filosofía que uno escribe para sus estudiantes de posgrado y sus colegas de la especialidad suele ser prácticamente ilegible, y por lo tanto casi nunca se lee.

Un curioso efecto secundario de mi política de tratar de escribir argumentos y explicaciones que fácilmente pueda entender la gente que no pertenezca a los departamentos de filosofía es que hay filósofos que por «principio» no se toman en serio mis argumentos. Cuando hace varios años dicté las John Locke Lectures en Oxford con auditorio lleno, se oyó a un distinguido filósofo salir de una de ellas refunfuñando que no estaba dispuesto a aprender nada de alguien que pudiera atraer a no filósofos a las Locke Lectures. Fiel a su palabra, nunca aprendió nada de mí, por lo que sé. No hice ajustes a mi estilo y nunca me he lamentado de pagar el precio. En la filosofía hay un momento y un lugar para los argumentos rigurosos, con todas las premisas numeradas y las reglas de inferencia expresadas, pero esto pocas veces necesita ostentarse en público. Les pedimos a nuestros estudiantes de posgrado que demuestren en sus tesis que pueden hacerlo, y desgraciadamente algunos nunca pierden la costumbre. Y para ser justos, el pecado opuesto, la rimbombante retórica continental, salpicada de ornamentos literarios e indicios de profundidad, tampoco le hace a la filosofía ningún favor. Si tuviera que elegir, me quedaría mil veces con el áspero analítico y sus hachazos de lógica, y no con el sabio

profundo y grandilocuente. Con el analítico normalmente puedes al menos hacerte una idea de lo que está diciendo y de lo que podría considerarse un error.

El punto medio, más o menos a mitad del camino entre la poesía y las matemáticas, es donde los filósofos pueden hacer sus mejores contribuciones, creo yo, y aportar genuinas aclaraciones de problemas profundamente enigmáticos. No hay algoritmos viables para hacer esta clase de trabajo. Como *todo* está disponible, uno escoge sus puntos fijos con la debida cautela. La mitad de las veces, una suposición «inocente» que todas las partes aceptaron sin previo aviso resulta ser la culpable. Para explorar estos peligrosos territorios conceptuales, las herramientas de pensamiento ideadas en el acto son de gran ayuda para aclarar los caminos alternativos y echar luz sobre sus posibilidades.

Estas herramientas de pensamiento casi nunca establecen un punto fijo *fijo* —un sólido «axioma» para toda investigación futura— sino que introducen un *candidato* digno de ser punto fijo, una probable limitación para la investigación futura, pero él mismo sujeto a revisión o a que de plano pueda tirarse por la borda si alguien resuelve por qué. No es de extrañar que a muchos científicos no les guste nada la filosofía; todo está disponible, nada es seguro sin asomo de duda, y las intrincadas redes de argumentos que se tejen para conectar estos puntos «fijos» cuelgan en el aire provisionalmente, sin estar atadas a los claros cimientos de la prueba empírica o la falsabilidad. Entonces, estos científicos le dan la espalda a la filosofía y siguen adelante con su trabajo, pero a costa de no tomar en cuenta algunas de las preguntas más importantes y fascinantes. «¡No preguntes! ¡No digas! ¡Es prematuro abordar el problema de la conciencia, del libre albedrío, de la moral, del significado y la creatividad!» Sin embargo, pocos pueden vivir con esa abstinencia, y en los últimos años, una especie de fiebre del oro ha traído a los científicos a estas apartadas regiones. Atraídos por una simple curiosidad (o a veces posiblemente por una búsqueda de fama), se embarcan en las grandes preguntas y pronto descubren lo difícil que es progresar en ellas. Debo confesar que uno de los deliciosos placeres, si bien culpables, que disfruto es ver a científicos eminentes que hace apenas unos años expresaban un desdén mordaz hacia la filosofía[7] dar traspiés vergonzosamente en sus propios empeños de aclarar estos asuntos con unas cuantas extrapolaciones de su propia investigación científica bruscamente argumentadas. Es todavía mejor cuando solicitan, y agradecen, un poco de ayuda de nosotros los filósofos.

En el siguiente capítulo presento 12 herramientas generales, para todo uso, y luego, en los subsecuentes, agrupo el resto de las entradas, no por tipo de herramienta, sino por el tema donde mejor funciona cada herramienta. Primero atiendo el tema filosófico más fundamental —el significado o contenido—, seguido de la evolución, la conciencia y el libre albedrío. Algunas de las herramientas que presento son como verdadero *software:* recursos sencillos que pueden hacer por tu imaginación lo que los telescopios y

microscopios pueden hacer por tu vista.

En el camino, también presentaré algunos falsos amigos, herramientas que echan humo en vez de alumbrar. Necesitaba un término para estos dispositivos peligrosos, y en mi experiencia náutica encontré *le mot juste*. Muchos navegantes se divierten con los términos náuticos que desconciertan a los marineros de agua dulce: babor y estribor, gorrón y pivote, obenques y estayes, jarcias y gateras, y todos los demás. En un barco en el que una vez navegué se hacía la broma de inventar definiciones falsas de estos términos. Así, *bitácora* era el nombre de las bioincrustaciones marinas en las brújulas, el *pescante* era un ceviche que se disfrutaba en la travesía, la *polea con gancho* era una maniobra de defensa personal y *pujamen* era vigor al trabajar. Desde entonces no he conseguido pensar en el pujamen —la orilla inferior de una vela— sin que me venga a la mente la imagen de unos marineros sudando del esfuerzo. Hay un accesorio desmontable que sirve para sostener la botavara cuando no está en uso, y tiene forma de muleta.[8] Inspirado en aquella experiencia y pensando en las partes de una embarcación, elegí llamar *muletas* a las herramientas de pensamiento que sólo *parecen* ayudar a entender pero que de hecho esparcen oscuridad y confusión en vez de luz. Como el artefacto ortopédico, si se nos caen, nos caemos con ellas. Hay, esparcidas a lo largo de estos apartados, una variedad de muletas de razonamiento, pero con las debidas notas de advertencia y ejemplos para deplorar. Cierro con algunas otras reflexiones sobre cómo es ser un filósofo, por si acaso alguien quiere saberlo, junto con algunos consejos del tío Dan para los lectores que puedan haberle encontrado el gusto a esta manera de investigar el mundo y se pregunten si están hechos para una carrera en este campo.

[1] Las palabras y frases en negritas son los nombres de las herramientas de pensamiento que se describen y analizan con más detalle en alguna otra parte del libro. Pueden encontrarse en el índice analítico, pues no todas tienen una sección entera dedicada a ellas.

[2] J. Searle, «Minds, Brains and Programs», *Behavioral and Brain Sciences,* 3 (3): 417-424, 1980, y D. C. Dennett, «The Milk of Human Intentionality», *Behavioral and Brain Sciences,* 3 (3): 428-430, 1980.

[3] Esopo, como Homero, es casi tan mítico como sus fábulas, que durante siglos se transmitieron oralmente antes de que se escribieran unos pocos cientos de años antes de la era de Platón y Sócrates. Es posible que Esopo no haya sido griego; hay pruebas circunstanciales de que era etíope.

[4] El filósofo W. V. O. Quine en *Word and Object* (The MIT Press, Cambridge, 1960) llamó a esto *ascenso semántico,* que va hacia *arriba:* de hablar acerca de electrones, justicia, caballos o lo que sea pasa a hablar acerca de *hablar acerca de* electrones, justicia, caballos o lo que sea. A veces la gente se opone a esta movida de los filósofos («¡Con ustedes todo es semántica!»), y a veces la movida es de hecho inútil o incluso embaucadora, pero cuando se necesita, cuando en una conversación la gente no habla de lo mismo o se engaña por las suposiciones tácitas acerca de lo que significan sus propias palabras, el ascenso semántico, o ponerse *meta-,* es la llave a la claridad.

[5] D. Hofstadter, *I Am a Strange Loop,* Basic Books, Nueva York, 2007.

[6] Muchos de los pasajes de este libro se han tomado de libros y artículos que publiqué antes, y se han revisado para hacerlos más portátiles y versátiles, apropiados para contextos distintos del original, que es una característica de la mayor parte de las buenas herramientas. Por ejemplo, la historia acerca de Von Neumann con la que abro estas páginas apareció en mi libro de 1995 *La peligrosa idea de Darwin: evolución y significados de la vida,* y la discusión sobre las herramientas manuales de Hofstadter apareció en mi artículo «Darwin's "Strange Inversion of Reasoning"» [«La "extraña inversión de razonamiento" de Darwin»], publicado en *Proceedings of the Natural Academy of Sciences* en 2009. En vez de consignarlos todos en notas a pie de página, al final del libro presento un listado de fuentes.

[7] Dos de los mejores: «La filosofía es a la ciencia lo que las palomas son a las estatuas», y «La filosofía es a la ciencia lo que la pornografía a la sexualidad: es más barata, más fácil, y algunas personas la prefieren» (no daré el crédito de estas frases, pero sus autores pueden reivindicarlas si lo desean).

[8] *Boom crutch,* literalmente «muleta de botavara». [t.]

II. Doce herramientas de pensamiento generales

La mayoría de las herramientas de pensamiento que presento en este libro son totalmente especializadas, hechas por encargo para aplicarse a un tema específico e incluso a una controversia específica dentro del tema. Sin embargo, antes de pasar a estas bombas de intuición, he aquí unas cuantas herramientas de pensamiento para todo uso, ideas y prácticas que han demostrado su utilidad en una amplia variedad de contextos.

1. Cometer errores

> El que dice «¡Más vale vivir siempre sin creer que creer una mentira!» simplemente muestra su propio miedo a que lo embauquen […] Es como un general diciendo a su tropa que es preferible no entrar nunca en combate a correr el riesgo de ser herido. Así no se vence ni al enemigo ni a la naturaleza. Nuestros errores no pueden ser esas cosas tan horriblemente fúnebres. En un mundo en el que es inevitable incurrir en ellos a pesar de toda nuestra cautela, cierta ligereza de ánimo parece más saludable que este desmesurado nerviosismo en su representación.
>
> WILLIAM JAMES, «The Will to Believe»

> Si uno está decidido a verificar una teoría, o si se desea explicar una cierta idea, en todos los casos debería publicarla, sea cual sea la forma en que resulte. Si solamente publicamos resultados de un cierto tipo, podemos hacer que los argumentos suenen bien. No, es preciso publicar *ambos* tipos de resultados.
>
> RICHARD FEYNMAN, *¿Está usted de broma, señor Feynman?*[1]

Los científicos a menudo me preguntan por qué los filósofos dedican tantos esfuerzos a enseñar y aprender la historia de su especialidad. Normalmente los químicos se las arreglan con apenas un conocimiento rudimentario de la historia de la química, que han aprendido sobre la marcha, y parece que no muchos biólogos moleculares tienen la menor curiosidad sobre lo que pasó en la biología antes de 1950. Mi respuesta es que la historia de la filosofía es en gran medida la historia de gente muy lista que comete errores muy tentadores, y que si no conoces la historia, estás condenado a volver a cometer los mismos malditos errores. Por eso enseñamos la historia del ramo a nuestros estudiantes, y los científicos que ignoran alegremente la filosofía lo hacen bajo su propio riesgo. No existe la ciencia libre de filosofía: sólo la ciencia que se ha llevado a cabo sin ninguna consideración de sus supuestos filosóficos subyacentes. Los científicos más listos o más suertudos a veces consiguen evitar los escollos con gran habilidad (a lo mejor ellos son «filósofos natos», o son tan listos como ellos mismos creen), pero son excepciones. No estoy diciendo que los filósofos profesionales no cometan también los viejos errores y a veces hasta los defiendan. Si las preguntas no fueran díficiles, no valdría la pena trabajar sobre ellas.

A veces uno no sólo quiere *arriesgarse* a cometer errores, sino en verdad cometerlos, aunque sólo sea para obtener algo claro y detallado que arreglar. Cometer errores es la clave para progresar. Desde luego, en ocasiones es verdaderamente importante no cometer ningún error, como cualquier cirujano o piloto aviador nos lo podrá confirmar. En cambio, no es tan ampliamente reconocido que también en ocasiones cometer errores es la única manera de avanzar. Muchos de los estudiantes que llegan a universidades muy competitivas se enorgullecen de no cometer errores: después de todo, así es como han llegado tanto más lejos que sus compañeros, o eso les han hecho creer. A menudo me doy cuenta de que tengo que alentarlos a *cultivar el hábito* de cometer errores, que son

las mejores oportunidades de aprender. Les entra el «bloqueo ante la página en blanco» y pasan horas dando vueltas desesperadamente de un lado al otro de la primera línea. «¡Desembúchalo!», los conmino. Y entonces ya tienen en la página algo con lo que pueden trabajar.

Los filósofos somos especialistas en el error (ya sé que suena a mala broma, pero déjenme terminar). Mientras que otras disciplinas se especializan en obtener las respuestas correctas a las preguntas que las definen, los filósofos nos especializamos en las diferentes maneras posibles de confundir y equivocar las cosas a tal punto que nadie está siquiera seguro de cuáles son las *preguntas* correctas, ya no se diga las respuestas. Si plantea las preguntas equivocadas, uno corre el riesgo de empezar cualquier investigación con el pie izquierdo. Siempre que eso pasa, ¡esto es un trabajo para filósofos! La filosofía, en todos los campos de investigación, es lo que tienes que hacer mientras averiguas cuáles son las preguntas que deberías haber planteado desde un principio. A algunos les molesta que eso pase. Preferirían tomar sus preguntas del perchero, todas muy bien confeccionadas, limpias y planchadas, y listas para responderse. Los que se sienten así pueden hacer física, matemáticas, historia o biología. Hay suficiente trabajo para todos. A los filósofos nos gusta trabajar en las preguntas que necesitan estirarse antes de que puedan ser respondidas. No es para todos. De todas formas haz un intento: a lo mejor te gusta.

A lo largo de este libro voy a brincar enérgicamente sobre lo que en mi opinión son errores de otras personas, pero téngase por seguro que yo mismo soy alguien con amplia experiencia en cometer errores. Tengo en mi haber algunos sobresalientes, y espero conseguir más. Uno de los propósitos de este libro es ayudarte a cometer *buenos* errores, de esos que alumbran el camino para todos.

Primero la teoría y después la práctica. Los errores no sólo son oportunidades de aprender; en un sentido importante, son la *única* oportunidad de aprender o de hacer algo verdaderamente nuevo. Antes de que haya aprendizaje debe haber aprendedores. Sólo hay dos maneras no milagrosas de que vengan al mundo los aprendedores: o a través de la evolución o porque los diseñaron y construyeron aprendedores que evolucionaron. La evolución biológica avanza mediante un enorme e inexorable proceso de ensayo y error, y sin los *errores* los ensayos no llevarían a ninguna parte. Como dijo Gore Vidal en una ocasión: «No basta con triunfar; otros tienen que fracasar». Los ensayos pueden ser o *ciegos* o *previsores.* Tú, lector, que conoces muchas cosas pero no la respuesta a la pregunta que nos ocupa, puedes dar saltos: saltos previsores. Puedes mirar antes de saltar, y por consiguiente tener de entrada *cierta* orientación a partir de lo que ya sabes. *Tú* no tienes que hacer cálculos aleatorios, pero tampoco desprecies los cálculos aleatorios. Entre los productos maravillosos de la evolución biológica estás tú.

La evolución es uno de los temas principales de este libro, como de todos mis libros,

por la sencilla razón de que es el proceso instrumental fundamental no sólo de la vida sino del conocimiento, el aprendizaje y el entendimiento. Si intentas explicarte el mundo de las ideas y los significados, el libre albedrío y la moral, el arte y la ciencia, e incluso la misma filosofía, sin un conocimiento sólido y detallado de la evolución, tienes una mano atada a la espalda. Más adelante examinaremos algunas herramientas concebidas para ayudarte a pensar en algunas de las preguntas sobre la evolución que más invitan a reflexionar, pero aquí necesitamos poner unos cimientos. Para que la evolución, que no sabe nada, llegue a algo novedoso, las mutaciones, que son «errores» aleatoriamente copiados en el ADN, tienen que avanzar a ciegas. La mayoría de estos errores tipográficos no tienen ninguna importancia, pues ¡nada los lee! Son tan intrascendentes como los borradores que nunca le entregaste al profesor para tu calificación. El ADN de una especie es como una receta para fabricar un nuevo cuerpo, y la mayor parte del ADN en realidad no se consulta durante el proceso de fabricación (a menudo se le llama «ADN basura» precisamente por esa razón). En las secuencias de ADN que sí se leen y se obedecen durante el desarrollo, la gran mayoría de las mutaciones son dañinas; de hecho, muchas de ellas causan la muerte rápidamente. Como la mayoría de las mutaciones «expresadas» son perjudiciales, el proceso de la selección natural en realidad trabaja para mantener muy bajo el índice de mutaciones. Cada uno de los lectores tiene un muy buen mecanismo copiante en sus células. Tú, por ejemplo, tienes aproximadamente un billón de células en el cuerpo, y cada una de ellas tiene una copia perfecta o *casi* perfecta de tu genoma, que tiene más de tres mil millones de símbolos y es la receta de ti que vio la luz en el momento en que el óvulo y el esperma de tus progenitores unieron fuerzas. Afortunadamente, el mecanismo copiante no consigue triunfos perfectos, pues si así fuera, la evolución tarde o temprano se detendría en seco, cuando se hubieran agotado sus fuentes de novedad. Esas manchitas, esas «imperfecciones» en el proceso, son la fuente de toda la asombrosa complejidad y de las maravillosas creaciones del mundo viviente (no me resisto a añadir: si algo merece llamarse *pecado original* son estos errores de copiado).

El secreto para cometer buenos errores es no ocultarlos: sobre todo no ocultártelos a ti mismo. En lugar de apartarte en actitud de negación cuando cometes uno, deberías volverte un conocedor de tus propios errores; darles vueltas en la cabeza como si fueran obras de arte —que en un sentido lo son—. La reacción fundamental a cualquier error debería ser: «¡Bueno, pues: no volveré a hacer *eso*!» En realidad, la selección natural no piensa en el pensamiento: simplemente borra las metidas de pata antes de que puedan reproducirse. La selección natural no volverá a hacer *eso,* al menos no con la misma frecuencia. Los animales que pueden aprender (aprender a no hacer ese ruido, tocar ese cable, probar esa comida) tienen *algo* con una fuerza selectiva similar en sus cerebros (B. F. Skinner y los conductistas comprendieron esto y lo llamaron aprendizaje por

«refuerzo»; *esa* respuesta no se refuerza y se «extingue»). Los seres humanos llevamos el asunto a un nivel mucho más veloz y eficiente. De hecho podemos *pensar en el pensamiento,* reflexionar sobre lo que acabamos de hacer: «¡Vaya: no volveré a hacer *eso*!» Y cuando reflexionamos, enfrentamos directamente el problema que toda persona que comete errores debe solucionar: ¿exactamente qué es *eso*? ¿Qué de lo que acabo de hacer me metió en tantos problemas? El secreto es aprovechar los detalles específicos del lío que has provocado, de modo que tu siguiente intento tome en cuenta esa información y no sean palos de ciego.

Todos hemos oído frases un poco desesperadas del tipo «Bueno, en ese momento no parecía tan mala idea». Dichos así pueden reflejar la reflexión arrepentida de un tonto, una señal de estupidez, pero lo cierto es que deberían considerarse pilares de sabiduría. Cualquier ser, cualquier agente que pueda en verdad decir «Bueno, en ese momento no parecía tan mala idea» está parado en el umbral de la inteligencia. Los seres humanos nos enorgullecemos de nuestra inteligencia, y uno de sus sellos distintivos es que podemos *recordar nuestro pensamiento previo* y reflexionar sobre él: sobre cómo era, por qué al principio parecía tentador, y qué se estropeó al final. No conozco indicios de que ninguna otra especie del planeta pueda pensar este pensamiento. Si así fuera, sería casi tan lista como la humana.

Así, pues, cuando cometas un error debes respirar hondo, aguantarte y *repasar;* sin misericordia ni apasionamiento, tu recuerdo del error. No es fácil. La reacción humana natural cuando se comete un error es avergonzarse y enojarse (cuando más nos enojamos es cuando nos enojamos con nosotros mismos), y tienes que hacer un gran esfuerzo para superar estas reacciones emocionales. Trata de adquirir la extraña práctica de saborear tus errores, deleitarte en sacar a la luz los extraños azares que te han llevado por el mal camino. Luego, cuando hayas succionado todo lo bueno que puedas obtener del hecho de haberlos cometido, podrás alegremente dejarlos tras de ti e ir en pos de la siguiente gran oportunidad. Sin embargo, con eso no basta: activamente debes buscar oportunidades de cometer grandes errores, para que entonces puedas recuperarte de ellos.

En su forma más simple, todos aprendimos esta técnica en la primaria. Recuerda cuán extraña y difícil parecía la división larga al principio: te enfrentabas a dos números imponderablemente largos, y tenías que calcular cómo empezar. ¿Cuántas veces cabe el divisor en el dividendo? ¿Siete, ocho? ¡Quién sabe! No tenías que saberlo: sólo tenías que hacer un intento, con el número que quisieras, y verificar el resultado. Recuerdo que me horroricé cuando me dijeron que debía empezar por calcular casi «adivinando». ¿No se trataba de *matemáticas*? Algo tan serio no era para estar jugando a las adivinanzas, ¿o sí? Con todo, a la larga aprecié, como todos, el encanto de la táctica. Si el número elegido resultaba ser demasiado pequeño, lo aumentabas y volvías a comenzar; si era

demasiado grande, lo disminuías. Lo bueno de la división larga era que siempre funcionaba, aunque tu primera elección hubiera sido la más tonta. En tal caso, solamente te tardabas un poco más.

Se han encontrado muchas aplicaciones para esta técnica general de hacer conjeturas con cierta base, calcular sus implicaciones y a partir del resultado hacer una corrección para la siguiente fase. Antes de que se inventara el GPS, los navegantes solían determinar su posición en el mar conjeturando dónde estaban (conjeturaban *exactamente* cuál sería su latitud y su longitud), y luego calculaban exactamente a qué altura del cielo aparecería el sol si por alguna increíble coincidencia ésa fuera su posición real. Cuando empleaban este método, no esperaban dar en el clavo. No tenían que hacerlo. Más bien, procedían a medir el ángulo de elevación real del sol (con exactitud) y comparaban los dos valores. Con unos pocos cálculos insignificantes más, esto les indicaba el tamaño de la corrección que tendrían que hacer a su conjetura inicial, y en qué dirección.[2] En este método es útil tener una muy buena conjetura la primera vez, pero no importa que esté destinada a ser un error; el chiste es cometer el error con lujo de detalles, a fin de que haya algo serio que corregir (los aparatos de GPS usan la misma estrategia de conjeturar y corregir para ubicar su posición relativa a los satélites).

Desde luego, mientras más complejo sea el problema que se te plantea, más difícil será el análisis. Los especialistas en inteligencia artificial llaman a esto el problema de «asignación de abono» (igual podría llamarse «asignación de culpa»). Resolver a qué abonar y qué culpar es uno de los problemas más espinosos de la inteligencia artificial, y también a la selección natural se le plantea. Todos los organismos de la tierra mueren tarde o temprano al cabo de una historia de vida más o menos complicada. ¿Cómo diablos podría la selección natural *ver a través* de la niebla de todos estos detalles a fin de resolver qué factores positivos «premiar» con descendencia y qué factores negativos «castigar» no dándoles hijos? ¿Será cierto que algunos de los hermanos de nuestros ancestros murieron sin haber tenido hijos *porque sus párpados tenían la forma equivocada*? Si no, ¿cómo podría el proceso de la selección natural explicar por qué nuestros párpados llegaron a tener las formas excelentes que tienen? Parte de la respuesta nos es conocida: siguiendo el viejo adagio «Si no está roto, no lo arregles», deja casi todas tus viejas y conservadoras soluciones de diseño en su lugar y corre tus riesgos, pero antes pon una red de seguridad. La selección natural conserva automáticamente *todo el terreno que ha preparado hasta ahora,* e intrépidamente explora grandes y pequeñas innovaciones; las grandes casi siempre conducen inmediatamente a la muerte. Es un terrible desperdicio, pero nadie lleva la cuenta. Nuestros párpados estaban ya en su mayor parte diseñados por la selección natural mucho antes de que hubiera seres humanos, primates o siquiera mamíferos. Han tenido más de cien millones de años para alcanzar la forma que tienen hoy día, con sólo unos pequeños retoques en los últimos

seis millones de años, desde nuestro ancestro en común con los chimpancés y los bonobos. Otra parte de la respuesta es que la selección natural trabaja con grandes cantidades de casos, en los que incluso las ventajas minúsculas se manifiestan en la estadística y pueden acumularse automáticamente (otras partes de la respuesta son detalles técnicos que no tienen lugar en esta discusión básica).

He aquí una técnica que quienes hacen trucos de magia con cartas (al menos los mejores) explotan con resultados sorprendentes (creo que no provocaré la ira de los magos al revelar el siguiente truco, pues no se trata de un truco *específico* sino de un profundo principio general). Un buen mago conoce muchos trucos de cartas que dependen de la suerte: no siempre funcionan, y ni siquiera funcionan con frecuencia. Hay algunos efectos —difícilmente podrían llamarse trucos— que podrían funcionar ¡sólo una de mil veces! Esto es lo que hay que hacer: empiezas por decirle al público que vas a hacer *un* truco, y sin decirles qué truco es, echas mano del efecto de uno en mil. Por supuesto, casi nunca funciona, así que pasas discretamente hacia el segundo intento —de un efecto que funcione aproximadamente una de cien veces, quizá—, y cuando también ése falle (y así será casi siempre), discretamente pasas al efecto número tres, que funciona sólo como una de 10 veces, así que más te vale estar preparado con el efecto número cuatro, que funciona, digamos, la mitad de las veces. Si nada de esto funciona (y para entonces *lo normal* sería que una de las redes de seguridad te hubiera salvado de éste que es el peor de los casos), tienes un efecto a prueba de errores, que no impresionará mucho a la multitud pero al menos es un truco infalible. Sería verdaderamente muy mala suerte que a lo largo de toda una función tuvieras que recurrir siempre a tu última red de seguridad, pero cada vez que consigas alguno de los efectos de más altos vuelos el público estará estupefacto. «¡Imposible! ¿Cómo demonios podrías haber sabido cuál era mi carta?» *¡Ajá!* No lo sabías, pero tenías una linda manera de hacer un esperanzado intento a ciegas que resultó. Al ocultarles a los espectadores todos los casos de «error» —los ensayos que no salieron—, creaste un «milagro».

La evolución funciona igual: los errores bobos tienden a ser invisibles, así que todo lo que vemos es una estupenda cadena de triunfos. Por ejemplo, la gran mayoría —mucho más de 90 por ciento— de las criaturas que jamás hayan vivido murieron sin haber tenido hijos, pero *ni uno solo de tus ancestros* tuvo ese destino. ¿Es o no es tener suerte en la vida?

Una gran diferencia entre hacer ciencia y hacer magia en el escenario es que mientras los magos ocultan sus intentos fallidos lo mejor que pueden, en la ciencia los haces en público. Haces alarde de tus errores, a fin de que todo el mundo pueda aprender de ellos. Así te beneficias de la experiencia de todos los demás, y no sólo de tu propio camino idiosincrásico a lo largo del mundo de los errores (es bien conocido cómo el físico Wolfgang Pauli expresó su desdén por el trabajo de un colega diciendo que «ni siquiera

estaba equivocado»; una falsedad clara que expones a los críticos es mejor que una papilla informe). Esto, por cierto, es otra razón por la que los humanos somos más listos que todas las demás especies. No es tanto que nuestros cerebros sean más grandes o más potentes, ni siquiera que tengamos el don de reflexionar sobre nuestros errores pasados, sino que *damos a conocer los beneficios* que nuestros cerebros individuales han obtenido a través de sus historias individuales de ensayo y error.[3]

Me llena de asombro cuántas personas realmente listas no entienden que pueden cometer grandes errores en público y salir indemnes. Conozco a investigadores distinguidos que hacen un poco el ridículo desviviéndose por evitar la necesidad de admitir que se equivocaron en algo. Por lo visto, nunca se han dado cuenta de que la tierra no se traga a la gente que dice «¡Uy!, tienes razón. Supongo que me equivoqué». De hecho, la gente disfruta cuando alguien reconoce que cometió un error. A todo tipo de gente le encanta señalar errores. La gente de espíritu generoso aprecia que le des la oportunidad de ayudar y que reconozcas cuando lo consigue; la gente de espíritu mezquino disfruta ponerte en evidencia. ¡Déjalos! De cualquier manera todos salimos ganando.

Desde luego, por lo general la gente no disfruta corregir los errores *estúpidos* de los demás. Debes tener algo que *valga la pena* corregir, algo en lo que sea original acertar o equivocarse, algo que requiera construir la clase de pirámide de pensamiento arriesgado que vimos en el ejemplo de los trucos con cartas. Si edificas cuidadosamente sobre el trabajo de los demás, puedes aventurarte y en una de ésas llegar a construir tu propio piso en voladizo. Y luego hay una bonificación sorpresa: si eres uno de los grandes corredores de riesgos, la gente se va a deleitar al corregir tus ocasionales errores estúpidos, que demuestran que no eres tan especial, que eres un metepatas común y corriente, igual que el resto de los mortales. Conozco a filósofos extremadamente cuidadosos que —al parecer— nunca han cometido un error en su trabajo. No tienden a obtener gran cosa, pero lo poco que producen es impecable, si bien no osado. Su especialidad es señalar los errores de los demás, y eso puede ser un servicio valioso, pero nadie excusa *sus* pequeños errores con una risita cordial. Es justo decir que, lamentablemente, su mejor obra a menudo se ve eclipsada y cae en el olvido, ahogada en el ruido de los carros que pasan con pensadores más audaces al volante. En el apartado 76 veremos que la práctica generalmente buena de cometer errores audaces tiene también otros desafortunados efectos secundarios. Metaconsejo: ¡no te tomes *ningún* consejo demasiado en serio!

2. «Por parodia de razonamiento»: el uso de la *reductio ad absurdum*

La barreta de la investigación racional, la gran palanca que hace valer la consistencia, es la *reductio ad absurdum,* literalmente la reducción de un argumento al absurdo. Tomas la aseveración o conjetura de que se trate y ves si puedes extraer alguna contradicción (o sólo implicaciones ridículas). Si puedes, esa proposición tiene que descartarse o ser enviada de vuelta al taller para su remodelación. Lo hacemos todo el tiempo sin tomarnos la molestia de exponer la lógica subyacente: «Si eso es un oso, entonces los osos tienen cuernos» o «No llegará a tiempo a la cena a menos que pueda volar como Supermán». Cuando se trata de una controversia teórica peliaguda, la barreta se esgrime con energía, pero aquí es difícil establecer la diferencia entre hacer una crítica justa y refutar caricaturizando. ¿Puede ser de plano tan estúpido tu oponente como para creer la proposición que con unos hábiles movimientos acabas de reducir al absurdo? Una vez califiqué un trabajo de un estudiante donde había una palabra mal escrita que parecía una serendipia: en lugar de escribir *paridad* escribió *parodia,* y con eso dio lugar a la deliciosa frase «por parodia de razonamiento», un nombre que según yo les queda muy bien a los argumentos por reducción al absurdo legítimos, que son de lo más comunes en el turbulento mundo de la controversia científica y filosófica.

Hace unos años asistí en el Instituto Tecnológico de Massachusetts a un seminario sobre ciencia cognitiva guiado por el lingüista Noam Chomsky y el filósofo Jerry Fodor. Recuerdo que el público se reía con sus divertidísimas refutaciones de cognitivistas de otras partes que no merecían su aprobación. En ese entonces Roger Schank, el director del laboratorio de inteligencia artificial de la Universidad de Yale, era la bestia negra, y a juzgar por la versión de Chomsky, Schank tenía que ser un pedazo de idiota o algo así. Yo conocía muy bien a Roger y su trabajo, y aunque tenía mis propios desacuerdos con él, pensaba que la versión de Noam era prácticamente irreconocible, así que levanté la mano para decir que posiblemente no se daba cuenta de algunas sutilezas de la posición de Roger.

«¡Oh, no! —insistió Noam, riendo— ¡Esto es lo que sostiene!», y siguió adelante con su trabajo de demolición, para gran regocijo de los presentes. Al cabo de unos cuantos minutos volví a intervenir: «Tengo que reconocer que las opiniones que critican son simplemente ridículas —aquí Noam sonrió en señal de asentimiento—, pero entonces me gustaría saber por qué pierden su tiempo y el nuestro criticando toda esa basura». Fue un balde de agua fría muy efectivo.

¿Y qué decir de mis propias reducciones al absurdo de las opiniones de otros? ¿Han sido más justas? He aquí algunas para considerar. Decida el lector. En una conferencia que tuvo lugar en Venecia, el neurocientífico francés Jean-Pierre Changeux y yo discutimos con el neurocientífico sir John Eccles y el filósofo sir Karl Popper sobre el

cerebro y la conciencia. Changeux y yo éramos los materialistas (que como tales sostenemos que la mente *es* el cerebro), y Popper y Eccles los dualistas (que sostienen que una mente no es algo material como un cerebro, sino algún otro segundo tipo de entidad que interactúa con el cerebro). Eccles había ganado el premio Nobel muchos años antes por el descubrimiento de la sinapsis, el espacio microscópico entre las neuronas que las moléculas de glutamato y otros neurotransmisores y neuromoduladores atraviesan billones de veces al día. De acuerdo con Eccles, el cerebro era como un potente órgano tubular y los billones de sinapsis constituían los teclados. La mente inmaterial —el alma inmortal de acuerdo con Eccles, un católico devoto— tocaba las sinapsis fomentando de algún modo que los niveles cuánticos de las moléculas de glutamato se animaran. «Olvídense de toda esa discusión teórica de redes neuronales y demás; son puras tonterías intrascendentes —dijo—. ¡La mente está en el glutamato!» Cuando me tocó hablar, dije que quería estar seguro de haber comprendido su posición. Si la mente estaba en el glutamato y yo echaba un tazón de glutamato por el desagüe, ¿no sería eso un asesinato? «Bueno —respondió, un poco desconcertado—, sería muy difícil saberlo, ¿no es así?»[4]

Uno pensaría que sir John Eccles, el dualista católico, y Francis Crick, el materialista ateo, tendrían muy poco en común, aparte de sus premios Nobel. Al menos por un rato, sin embargo, sus respectivas ideas sobre la conciencia compartieron una discutible sobresimplificación. Muchos no científicos no se dan cuenta de que las sobresimplificaciones pueden ser maravillosas para la ciencia: permiten abrirse camino entre la espantosa complejidad con un prototipo que es *casi* correcto, y dejar los sucios detalles para después. Podría decirse que el mejor uso de la «sobre»-simplificación en la historia de la ciencia fue la táctica evasiva de Crick y James Watson para encontrar la estructura del ADN mientras Linus Pauling y otros avanzaban laboriosamente intentando entender todos los detalles. Crick estaba totalmente a favor de intentar la jugada atrevida sólo si resolvía el problema de un tirón, pero desde luego que esto no siempre funciona. Una vez tuve oportunidad de demostrarlo en una de las famosas tardes de té a las que convocaba Crick en La Jolla. Estas tertulias vespertinas eran reuniones de laboratorio informales, en las que los visitantes podían plantear problemas y participar en la discusión general. En aquella ocasión Crick hizo una declaración atrevida: recientemente se había demostrado que a las neuronas del área cortical V4 «les importaba» el color (respondían diferencialmente a él). Y entonces propuso una hipótesis sorprendentemente simple: la experiencia consciente del rojo, por ejemplo, *es* actividad en las neuronas competentes sensibles al rojo de esa área de la retina. «¡Um! —me asombré—. ¿Entonces estás diciendo que si quitáramos algunas de esas neuronas sensibles al rojo y las mantuviéramos vivas en una caja de Petri, y las estimuláramos con un microelectrodo, habría *conciencia de rojo* en la caja de Petri?» Una manera de

responder cuando te presentan una reducción al absurdo es coger el toro por los cuernos y aprobar la conclusión, una movida que una vez llamé *pasarse de listo,* en inglés *outsmarting ,* pues el filósofo australiano J. J. C. Smart era célebre por decir que sí, de acuerdo con su teoría de la ética, ¡a veces *sí* era correcto incriminar y colgar a un hombre inocente! Crick decidió «pasarse de listo» conmigo. «¡Sí! ¡Sería un caso aislado de conciencia de rojo!» ¿La conciencia de rojo de quién? Esto no lo dijo. Más adelante refinó su pensamiento a ese respecto, pero, con todo, el neurocientífico Christof Koch y él, en su búsqueda de lo que llamaban los correlatos neurales de la conciencia, nunca abandonaron por completo su lealtad a esta idea.

Tal vez otro encuentro realce más lo problemático de la idea de una pizca de conciencia en un recipiente. El físico y matemático Roger Penrose y el anestesiólogo Stuart Hameroff se asociaron para elaborar una teoría de la conciencia que dependiera, no del glutamato, sino de los efectos cuánticos en los microtúbulos de las neuronas (los microtúbulos son cadenas de proteínas tubulares que sirven como vigas y carreteras dentro del citoplasma de todas las células, no sólo las neuronas). En Tucson II, la segunda conferencia internacional sobre la ciencia de la conciencia, después de que Hameroff expusiera esta postura, pregunté desde las butacas: «Stuart, tú eres anestesiólogo. ¿Alguna vez has prestado tu ayuda en alguna de esas cirugías dramáticas en las que se remplaza alguna mano o un brazo amputado?» No lo había hecho, pero sabía de ellas. «Dime si hay algo que no esté entendiendo, Stuart, pero dada tu teoría, si fueras el anestesiólogo en una de esas operaciones te sentirías moralmente obligado a anestesiar la mano amputada que pusieron en hielo, ¿no te parece? Después de todo, los microtúbulos de los nervios de la mano estarían haciendo lo suyo, tal como los microtúbulos del resto del sistema nervioso, y esa mano estaría muy dolorida, ¿no es así?» La expresión de Stuart indicaba que esto nunca se le había ocurrido. La idea de que la conciencia (de rojo, de dolor, de lo que sea) es una especie de propiedad en red, algo que supone miríadas de neuronas en actividad coordinada, puede no ser muy atractiva de entrada, pero estos intentos de reducción al absurdo pueden ayudar a que la gente vea por qué debe tomarse en serio.

3. Las reglas de Rapoport

¿Qué tan benévolo se espera que seas cuando criticas las opiniones de un oponente? Si hay contradicciones *obvias* en lo que el oponente defiende, entonces por supuesto que debes señalarlas, y con energía. Si hay contradicciones más o menos ocultas, debes cuidadosamente sacarlas a la luz, y luego demolerlas. Sin embargo, la manía de buscar contradicciones ocultas puede fácilmente convertirse en una actitud tiquismiquis, de abogado naval[5] o, como hemos visto, en parodia declarada. La emoción de la caza y la convicción de que tu oponente *tiene* que albergar una confusión en algún lugar alienta la interpretación no benévola, que te deja un fácil blanco de ataque, pero esos blancos fáciles suelen ser irrelevantes para los temas reales que están en juego y simplemente hacen que todo el mundo pierda el tiempo y la paciencia, aunque diviertan a tus partidarios. El mejor antídoto que conozco contra esta tendencia a caricaturizar al oponente es una lista de reglas que hace muchos años promulgó el psicólogo social y teórico del juego Anatol Rapoport (creador de la estrategia del toma y daca ganadora en el legendario torneo del dilema del prisionero organizado por Robert Axelrod).[6]

Cómo escribir un comentario crítico exitoso:

1. Intenta reexpresar la posición del blanco de tu crítica con tal claridad, viveza e imparcialidad que él mismo diga «Gracias, me gustaría haberlo expresado así».
2. Haz una lista de todos sus puntos de acuerdo (especialmente si no se trata de acuerdos generales o extendidos).
3. Menciona cualquier cosa que hayas aprendido del blanco de tu crítica.
4. Sólo cuando lo hayas hecho estarás autorizado para decir aunque sea una sola palabra para refutar o criticar.

Cuando se siguen estas reglas, un efecto inmediato es que los blancos de tus críticas se vuelven más receptivos a ellas: ya has mostrado que comprendes su postura tan bien como ellos, y también has demostrado tener buen juicio (coincides con ellos en algunos asuntos importantes e incluso algo que han dicho te ha convencido).[7]

Al menos a mí me cuesta mucho seguir las reglas de Rapoport. Algunos blancos de nuestras críticas francamente no merecen esa atención respetuosa y —lo admito— ensartarlos y rostizarlos puede ser un verdadero gozo. Sin embargo, cuando se requiere, y funciona, los resultados son gratificantes. En mi libro *Freedom Evolves*[8] fui

especialmente esmerado en mi intento de hacer justicia al tipo de incompatibilismo que sostiene Robert Kane,[9] una postura frente al libre albedrío con la que estoy en profundo desacuerdo, y atesoro la respuesta que me escribió cuando le mandé el capítulo en borrador:

> De hecho me gusta mucho, no obstante nuestras diferencias. El tratamiento de mi postura es amplio y, en líneas generales, justo, mucho más que el que normalmente uno recibe de los críticos. Transmites la complejidad de mi postura y la seriedad de mis empeños por abordar preguntas difíciles en vez de evadirlas. Te estoy agradecido por esto y por haberle dedicado tantas páginas.

Otras personas que fueron objeto de mi atención guiada por Rapoport han sido menos cordiales. Mientras más justa parece la crítica, más difícil de sobrellevar es en algunos casos. Vale la pena recordar que un intento heroico por encontrar una interpretación defendible de un autor, si resulta infructuoso, puede ser incluso más aplastante que una crítica feroz. Lo recomiendo.

4. La ley de Sturgeon

El autor de ciencia ficción Ted Sturgeon, en una conferencia que pronunció en septiembre de 1953 en la Convención Mundial de Ciencia Ficción en Filadelfia, dijo:

> Cuando la gente habla de la novela de misterio, menciona *El halcón maltés* y *El gran sueño*. Cuando se habla del *western,* se señalan *Camino de Oregón* y *Shane, el desconocido*. Sin embargo, cuando hablan de ciencia ficción, la llaman «esas cosas tipo Buck Rogers» y dicen que «noventa por ciento de la ciencia ficción es basura». Pues bien, tienen razón. Noventa por ciento de la ciencia ficción es basura. De hecho, noventa por ciento de todo es basura. Es el diez por ciento que no es basura lo que importa, y el diez por ciento de la ciencia ficción que no es basura es tan bueno o mejor que cualquier cosa que se haya escrito.

Normalmente la ley de Sturgeon se expresa con un poco menos de decoro: *Noventa por ciento de todo es pura mierda.* Noventa por ciento de los experimentos de biología molecular, 90 por ciento de la poesía, 90 por ciento de los libros de filosofía, 90 por ciento de los artículos de matemáticas que se escriben para revistas arbitradas —y así sucesivamente— es pura mierda. ¿Es cierto eso? Bueno, tal vez sea una exageración, pero convengamos en que en todos los campos se hace un montón de trabajo mediocre (algunos cascarrabias dicen que más bien es cerca de 99 por ciento, pero no entremos en ese juego). Una buena moraleja que puede extraerse de esta observación es que cuando quieras criticar un campo, un género, una disciplina, una forma de arte… ¡no *pierdas tu tiempo, y el nuestro, abucheando a la mierda!* Lánzate contra lo bueno, o déjalo en paz. Suelen ignorar este consejo los ideólogos que se concentran en destruir la reputación de la filosofía analítica, la psicología evolucionista, la sociología, la antropología cultural, la macroeconomía, la cirugía plástica, el teatro de improvisación, las comedias televisivas, la teología filosófica, el masaje terapéutico, lo que se te ocurra. Estipulemos de entrada que hay toda clase de cosas deplorables, estúpidas y de segunda categoría, y en gran cantidad. Ahora bien, a fin de que no pierdas tu tiempo ni pongas a prueba nuestra paciencia, asegúrate de que te concentras en lo mejor que puedas encontrar, los ejemplos propios del buque insignia, los que exaltan los líderes del ramo, las anotaciones ganadoras, no la escoria. Observa que esto se relaciona de cerca con las reglas de Rapoport: a menos que seas un comediante cuyo propósito principal sea hacer a la gente reír con tus payasadas, ahórranos la caricatura. Me he dado cuenta de que esto es especialmente cierto cuando el blanco es un filósofo. Con un par de buenas puntadas, hasta las mejores teorías y análisis de *cualquier* filósofo, desde los más grandes e inteligentes sabios de la antigua Grecia hasta los héroes intelectuales del pasado reciente

(Bertrand Russell, Ludwig Wittgenstein, John Dewey, Jean-Paul Sartre, por mencionar a cuatro pensadores muy diferentes entre sí), pueden presentarse como si fueran una completa idiotez —o pura quisquillosidad fastidiosa—. ¡Puaj! No lo hagas. Al único que desacreditarás será a ti mismo.

5. La navaja de Ockham

Atribuida a Guillermo de Ockham (u Occam o también Ockam), el lógico y filósofo del siglo XIV, esta herramienta de pensamiento es de hecho una regla general mucho más antigua. Un nombre para ella en latín es *lex parsimoniae,* la ley de la parsimonia. Suele resumirse en la máxima de «no multiplicar las entidades sin necesidad». La idea es sencilla: no te inventes una teoría complicada y extravagante si tienes una más simple (con menos ingredientes, menos entidades) que trate el fenómeno igual de bien. Si la exposición a un aire sumamente frío puede explicar todos los síntomas del congelamiento, no postules la existencia de «gérmenes de la nieve» o «microbios árticos» que nadie ha visto. Las leyes de Kepler explican las órbitas de los planetas; no hay ninguna necesidad de plantear la hipótesis de que unos pilotos conducen los planetas desde unos tableros de control ocultos bajo la superficie. Hasta aquí no es polémico, pero algunas extensiones del principio no siempre han recibido aprobación.

Conwy Lloyd Morgan, un psicólogo británico del siglo XIX, extendió la idea para abarcar atribuciones de mentalidad a los animales. El canon de parsimonia de Lloyd Morgan nos aconseja no atribuir mentes elaboradas a los insectos, los peces, y ni siquiera a los delfines, los perros y los gatos, si su conducta puede explicarse en términos más simples: «En ninguna circunstancia deberá interpretarse una actividad animal desde el punto de vista de procesos psicológicos superiores si perfectamente puede interpretarse desde el punto de vista de procesos que están más abajo en la escala de la evolución y el desarrollo psicológicos».[10] Si se abusa de esto parecería que nos están dando la orden de tratar a todos los animales e incluso a los seres humanos como si tuvieran cerebros pero no mentes. Como veremos, las tensiones que surgen cuando el tema de que se trata son las mentes no se han terminado de resolver ni están sujetas a prohibiciones absolutas.

Uno de los intentos menos admirables de aplicar la navaja de Ockham a un problema complicado es la afirmación (y las consecuentes refutaciones) de que postular la existencia de un dios como creador del universo es más simple, más ahorrativo, que las alternativas. ¿Cómo podría ser ahorrativo postular algo sobrenatural e incomprensible? A mí me da la impresión de ser la cima del despilfarro, pero tal vez haya maneras más inteligentes de rebatir esa teoría. No quiero discutir sobre eso; a fin de cuentas, la navaja de Ockham es sólo una regla general, una sugerencia que muchas veces es útil. La perspectiva de convertirla en un Principio Metafísico o Requisito Fundamental de Racionalidad que pudiera cargar con el peso de demostrar la existencia o la no existencia de Dios de un tirón es sencillamente absurda. Sería como tratar de desmentir un teorema de la mecánica cuántica demostrando que contradice el axioma «No pongas todos los huevos en la misma canasta».

Algunos pensadores han llevado la navaja de Ockham a extremos drásticos y la usan para negar la existencia del tiempo, la materia, los números, los agujeros, los dólares, los

programas de cómputo, etc. Uno de los primeros pensadores ultramezquinos fue el filósofo de la Grecia antigua Parménides, cuyo catálogo de cosas existentes era en efecto mínimo. Como memorablemente escribió en un examen un estudiante que tuve, «Parménides es el que dijo "Sólo hay una cosa, y esa cosa no soy yo"». Lamento decirlo, pero eso sí parece ser lo que Parménides trataba de decirnos. Sin duda, algo se pierde en la traducción. Los filósofos estamos acostumbrados a tomarnos en serio esas ideas, aunque sólo sea porque nunca podemos adivinar cuándo una idea «loca» va a juzgarse de manera injusta y desacertada, víctima de falta de imaginación.

6. La escoba de Ockham

El biólogo molecular Sidney Brenner inventó recientemente un exquisito juego con la navaja de Ockham y propuso la expresión *escoba de Ockham* para describir el proceso en el que paladines intelectualmente deshonestos de una u otra teoría barren por debajo de la alfombra los hechos inconvenientes. He aquí nuestra primera *muleta,* una herramienta de *antipensamiento,* y debes mantener un ojo muy avizor para cuidarte de ella. La práctica es especialmente insidiosa en manos de propagandistas que dirigen sus esfuerzos al público lego, porque, como la famosa pista de Sherlock Holmes acerca del perro que *no* ladraba en la noche, nadie más que los expertos puede notar la *ausencia* de un hecho que se ha barrido de la escena con la escoba de Ockham. Por ejemplo, los creacionistas invariablemente dejan fuera las abundantes pruebas embarazosas de las que sus «teorías» no pueden encargarse, y para alguien que no sea biólogo sus explicaciones hábilmente elaboradas pueden ser de lo más convincentes simplemente porque el lector lego no puede ver lo que *no está ahí.*

¿Cómo diablos puede uno estar al acecho de algo invisible? Con la ayuda de los expertos. *Signature in the Cell,* de Stephen C. Meyer,[11] pretende exponer la sistemática imposibilidad de que la vida tenga un origen natural (no sobrenatural) y presenta lo que incluso a un lector bien informado puede parecerle una investigación imparcial y exhaustiva de los modelos y teorías con los que se trabaja en el mundo entero, para mostrar que todos ellos son irremediablemente inútiles. ¡Sus argumentos son tan convincentes que en noviembre de 2009 el eminente filósofo Thomas Nagel lo declaró su Mejor Libro del Año en el *Times Literary Supplement* de Londres, una de las publicaciones de crítica de libros más influyentes del globo![12] En una vehemente correspondencia que mantuve con él después de que apareció su alabanza, Nagel demostró que sabía mucho de la historia de la investigación sobre el origen de la vida, suficiente para pensar que podría confiar en su propio criterio. En una carta al *Times Literary Supplement* observó: «El libro de Meyer me parece escrito de buena fe».[13] Si Nagel hubiera consultado a científicos del ramo, habría podido ver cómo Meyer explota la escoba de Ockham y barre factores inconvenientes para no verlos, y a lo mejor también le habría consternado enterarse de que a los especialistas nadie les había enviado un ejemplar del libro de Meyer para reseña, como a él, ni les habían pedido que lo arbitraran antes de su publicación. Enterarse de que el libro que admiraba era una operación furtiva habría sacudido su confianza en su propio juicio, o tal vez no. El *establishment* científico tiene fama de aplastar injustamente, de vez en cuando, a los críticos renegados, y a lo mejor —a lo mejor— Meyer no tuvo más remedio que lanzar su ataque a hurtadillas. Sin embargo, habría sido prudente que Nagel explorara con cautela esta posibilidad antes de comprometerse. Es justo decir que los científicos que

investigan el origen de la vida aún no tienen una teoría segura con la que todos estén de acuerdo, pero no hay escasez de candidatos, al contrario: más que una palestra casi vacía, hay muchas de donde elegir.

Los teóricos de la conspiración son expertos en la escoba de Ockham. Un ejercicio instructivo que puede hacerse en internet es buscar una nueva teoría de la conspiración para ver si tú, que no eres experto en el tema, puedes encontrar las fallas, antes de buscar en algún otro sitio de la red las refutaciones de los expertos. Cuando Brenner acuñó el término no estaba hablando de creacionismo ni de teorías de la conspiración: estaba señalando que a veces, en el fragor de la batalla, ni siquiera los científicos serios se resisten a «pasar por alto» algunos datos que podrían debilitar seriamente su teoría predilecta. Es una tentación que hay que resistir a toda costa.

7. Usar público lego como señuelo

Una buena manera de evitar que la gente empuñe *sin darse cuenta* la escoba de Ockham es una técnica que llevo años recomendando y que varias veces he puesto a prueba, pero nunca tan ambiciosamente como quisiera. A diferencia de las otras prácticas que he descrito, para llevar a cabo ésta como es debido se necesita tiempo y dinero. Espero que otros ejerzan esta técnica vigorosamente e informen sobre los resultados. He decidido ponerla aquí porque toca algunos de los problemas de comunicación que también las otras herramientas generales enfrentan.

En muchos campos de estudio, no sólo la filosofía, hay controversias que parecerían interminables y susceptibles de conducir a errores de interpretación: como cada loco con su tema, la gente habla sin escuchar al otro y no hace el esfuerzo necesario para comunicar sus ideas eficazmente. Se enardecen los ánimos, empiezan las señales de falta de respeto y desdén. Los que se mantenían al margen empiezan a tomar partido, aunque no entiendan del todo el asunto.

Se puede poner feo, y puede ser que su causa sea muy sencilla. Cuando los expertos hablan con expertos, pertenezcan o no a la misma disciplina, siempre *explican de menos.* No hay que ir muy lejos para encontrar la razón: explicar de más al colega experto es un insulto muy serio —«¿Te lo tengo que deletrear?»—, y nadie quiere insultar al colega experto. Así, para no correr peligro, la gente peca de explicar de menos. La mayoría de las veces no se hace a propósito, y es casi imposible evitarlo —lo cual no está mal, de hecho, pues ser amable de manera natural es un lindo rasgo de carácter en quien sea—, pero esta disposición cortés de asumir mayor entendimiento de lo apropiado en el distinguido público que lo escucha a uno tiene un desafortunado efecto secundario: que los expertos no se entiendan entre sí.

No hay una cura *directa:* si a todos los expertos presentes en un taller o conferencia se les pide que no expliquen de menos sus posturas, seguramente todos lo van a acatar, pero no funcionará. Si acaso, empeorará las cosas, pues entonces la gente estará especialmente sensible al riesgo de insultar a alguien sin querer. Sin embargo, sí hay una cura indirecta y de lo más efectiva: poner a todos los expertos a presentar sus posturas a un pequeño público de no expertos curiosos (aquí en Tufts tengo la ventaja de que hay alumnos de licenciatura muy brillantes), mientras los otros expertos escuchan sin intervenir. No tienen que escuchar a escondidas; no estoy sugiriendo *tejemanejes.* Al contrario, todos deberían saber y estar completamente enterados de que el propósito del ejercicio es que los participantes se sientan cómodos hablando con términos que *todos* van a entender. Al dirigir sus comentarios a los estudiantes de licenciatura (el público señuelo), los oradores no tienen que preocuparse *en absoluto* de insultar a los expertos, pues no se están dirigiendo a ellos (supongo que podrían preocuparse de insultar a los estudiantes de licenciatura, pero eso es otro asunto). Cuando todo marcha bien, el

experto A explica los problemas de la controversia a los estudiantes de licenciatura mientras el experto B escucha. En algún momento puede ser que se le ilumine la cara al experto B. «¡Así que *eso* es lo que has estado tratando de decir! Ahora entiendo.» O a lo mejor los buenos efectos tendrán que esperar a que le toque a B explicar a los mismos estudiantes cuáles son los problemas, y provoque la misma reacción grata en A. Puede ser que no salga perfecto, pero normalmente sale bien y todo el mundo se beneficia. Los expertos disuelven algunos de los malentendidos entre sus posturas, producto de errores de observación, y los estudiantes de licenciatura viven una experiencia educativa de primera.

Varias veces he armado este tipo de ejercicios en Tufts, gracias a un generoso apoyo de la administración. Selecciono cuidadosamente a un pequeño grupo de estudiantes de licenciatura (no más de 12) y les doy instrucciones sobre su papel: no deben aceptar nada que no entiendan. Se espera que levanten la mano, interrumpan, alerten a los expertos cuando algo les parezca vago o confuso (sí se les dan algunas lecturas obligatorias para que estudien previamente y así no sean unos completos novatos en el tema, sino *amateurs* interesados). Les encanta ese papel, y hay una buena razón: reciben de unos peces gordos tutoriales a la medida. Los expertos, por su parte, a menudo se dan cuenta de que el hecho de que se les asigne (con mucha antelación) la tarea de explicar su posición bajo estas condiciones les ayuda a encontrar maneras de plantear sus ideas que jamás se les habían ocurrido. A veces estos expertos llevan años «protegidos» con capas de colegas expertos, posdoctorados y estudiantes avanzados de posgrado, y de verdad necesitan una incitación así.

8. Pensar afuera de la caja

Es difícil *encontrar* una aplicación de la escoba de Ockham, pues actúa barriendo hechos inconvenientes para no verlos, y es todavía más difícil conseguir lo que Doug Hofstadter llama en inglés *jootsing* , por «*j*umping *o*ut *o*f *t*he *s*ystem», que en español es literalmente «bajarse del sistema», algo así como pensar afuera de la caja y salirse de lo convencional.[14] Es una táctica importante, no sólo en ciencia y en filosofía, sino en las artes. La creatividad, esa virtud que se busca fervientemente pero rara vez se encuentra, a menudo es una transgresión, que hasta el momento nadie ha imaginado, de las reglas del sistema del que surge. Puede ser el sistema de la armonía clásica en la música, las reglas de la métrica y el ritmo en los sonetos (hasta en los epigramas) o los «cánones» del gusto o la propiedad en algún género artístico. También puede ser las suposiciones y principios de alguna teoría o programa de investigación. Para ser creativo no basta con tratar de encontrar algo original —cualquiera puede hacer eso, pues puede encontrarse originalidad en cualquier yuxtaposición aleatoria de cosas—, sino que hay que salirse de algún sistema (o de algún cuadro), un sistema que por buenas razones ha llegado a ser algo establecido. Cuando una tradición artística alcanza el punto en el que literalmente «todo se vale», los que quieren ser creativos tienen un problema: no hay reglas fijas contra las cuales rebelarse, no hay expectativas complacientes que hacer añicos, nada que subvertir, ningún fondo sobre el cual crear algo que sea sorprendente pero significativo. *Conocer* la tradición ayuda si quieres subvertirla. Por eso a tan pocos diletantes o novatos se les ocurre nada verdaderamente creativo.

Siéntate a un piano y trata de inventar una buena nueva melodía; muy pronto verás lo difícil que es. Puedes usar todas las teclas, en cualquier combinación que elijas, pero mientras no encuentres algo en que apoyarte, mientras no definas algún estilo, género o motivo que explotar un poco, o al cual aludir, antes de torcerlo, no saldrá nada más que ruido. Y no cualquier transgresión funciona. Sé que hay al menos dos arpistas de jazz a los que les va bien (bueno, que sobreviven), pero pretender hacerse de un nombre interpretando a Beethoven en bongós tal vez no sea la mejor idea. Es aquí donde el arte comparte una característica con la ciencia: en cualquier altercado teórico siempre habrá montones de suposiciones sin examinar, pero tratar de negarlas una por una hasta encontrar una vulnerable no es una buena receta para el éxito en ciencia ni en filosofía (sería como tratar de tomar una melodía de Gershwin y alterarla nota por nota, en busca de un descendiente digno de ella. ¡Buena suerte! Las mutaciones casi siempre son nocivas). Es más difícil que eso, pero a veces se corre con suerte.

Sugerirle a alguien que progrese pensando afuera del cuadro es como sugerirle a un inversionista que compre barato y venda caro. Sí, desde luego, ésa es la idea, ¿pero cómo lo consigues? Obsérvese que la sugerencia de inversión no es *enteramente* vacua o inservible, y el llamado a pensar afuera de la caja es todavía más útil, pues aclara qué

aspecto tiene tu objetivo, por si algún día alcanzas a verlo (todo el mundo sabe qué aspecto tiene *más dinero*). Cuando te enfrentas a un problema científico o filosófico, el sistema del que necesitas salirte está normalmente tan afianzado que es tan invisible como el aire que respiras. Como regla general, cuando una controversia que lleva muchos años no parece llevar a ningún lado y las dos «partes» tercamente insisten en tener la razón, la mitad de las veces el problema es que hay algo en lo que ambos están de acuerdo y que no es cierto. Ambas partes lo consideran tan obvio, de hecho, que se da por sentado. Encontrar estos envenenadores invisibles de problemas no es tarea fácil, pues lo que parece obvio a estos expertos enfrentados es propenso a parecerle obvio, pensándolo bien, prácticamente a cualquiera. Así, es poco probable que la recomendación de estar al acecho de una falsa suposición tácita compartida rinda frutos, pero al menos es más probable que encuentres una si esperas hacerlo y tienes alguna idea de cuál sería su aspecto.

A veces hay pistas. Varios de los grandes ejemplos de pensamiento fuera del cuadro han significado abandonar alguna *cosa* que se tenía en gran estima y que resultó no existir a fin de cuentas. Se suponía que el *flogisto* era un elemento del fuego, y que el *calórico* era el fluido o gas invisible y autorrepelente que era el *ingrediente principal del calor, pero ambos se abandonaron, igual que el éter* como medio en el cual la luz viajaba tal como el sonido viaja a través del aire y del agua. Otros admirables ejemplos de pensamiento fuera del cuadro, empero, son sumas, no restas: los gérmenes, los electrones ¡y, tal vez, incluso los universos paralelos de la mecánica cuántica! Nunca es obvio desde el principio si deberíamos pensar afuera de la caja o no. Ray Jackendoff y yo hemos sostenido que es necesario abandonar la casi siempre tácita suposición de que la conciencia es el fenómeno mental «más elevado» o «fundamental», y yo he argumentado que pensar en la conciencia como un *medio* especial (un poco como el éter) en el que los contenidos se «transducen» o traducen es un hábito de pensamiento muy extendido del que se debe salir. Como muchos otros, también yo he argumentado que si piensas que es evidente que el libre albedrío y el determinismo son incompatibles, estás en un grave error. Más adelante abundaré sobre estas ideas.

Otra pista: a veces un problema comienza porque alguien en un pasado remoto dijo «Supongamos, por mor del argumento, que…», y la gente estuvo de acuerdo, por mor del argumento, y entonces, en las posteriores estocadas y evasiones, ¡a todo el mundo se le olvida cómo empezó el problema! Pienso que en ocasiones, al menos en mi especialidad dentro de la filosofía, los oponentes disfrutan tanto la pelea que ninguna parte quiere arriesgarse a que se cancele todo el ejercicio por examinar las premisas que lo hacen posible. He aquí dos viejos ejemplos, que desde luego son controvertidos: *1.* «¿Por qué hay algo y no más bien nada?» es una pregunta profunda que necesita respuesta. *2.* «¿Dios ordena algo porque es bueno, o algo es bueno porque Dios lo

ordena?» es otra pregunta importante. Supongo que sería maravilloso si a alguien se le ocurriera una *buena* respuesta a cualquiera de estas preguntas; por eso reconozco que llamarlos seudoproblemas a los que no vale la pena que nadie les preste ninguna atención no es muy agradable, pero eso no demuestra que esté equivocado. Nadie dijo que la verdad tenía que ser divertida.

9. Tres especies de *gouldeo:* masbienismo, apilamiento y el pasodoble de Gould

El biólogo Stephen Jay Gould fue un virtuoso diseñador y explotador de muletas de razonamiento. Aquí tenemos tres especies emparentadas entre sí, del género *Gouldae:* así las bauticé en honor de la persona que con más efectividad las ha manejado.

El *masbienismo* es una manera rápida y discreta de hacerte pasar de largo por una *falsa dicotomía.* La forma general del masbienismo es «No es cierto que *bla bla bla,* como te quiere hacer creer la ortodoxia; es *más bien* que *tal y tal y tal,* que es radicalmente diferente». Algunos masbienismos están muy bien; en verdad tienes que escoger entre las dos alternativas que se te ofrecen; en estos casos, no se te plantea una alternativa falsa, sino más bien una genuina e ineludible. Sin embargo, algunos masbienismos son poco más que prestidigitación, dado el hecho de que la expresión «más bien» implica —sin argumentos— que hay una importante incompatibilidad entre las afirmaciones que están a sus costados.

He aquí un buen ejemplo de masbienismo que comete Gould durante su explicación del equilibrio puntuado:

> El cambio normalmente no ocurre por una alteración imperceptiblemente gradual de especies enteras, sino *más bien* [las cursivas son mías] por aislamiento de pequeñas poblaciones y su geológicamente instantánea transformación en nuevas especies.[15]

Este pasaje nos invita a creer que el cambio evolutivo no podría ser al mismo tiempo «geológicamente instantáneo» e «imperceptiblemente gradual», pero desde luego que sí puede serlo. De hecho, es justamente eso lo que el cambio evolutivo debe ser, a menos que Gould esté diciendo que la evolución tiende a avanzar por saltaciones (saltos gigantes en el espacio de diseño); pero en otro lugar ha insistido en que nunca ha estado de acuerdo con el saltacionismo. La especiación «geológicamente instantánea» puede ocurrir a lo largo de un «corto» periodo: digamos, 50 000 años, un transcurso de tiempo que apenas si se detecta en la mayoría de los estratos geológicos. Durante ese breve momento, el miembro típico de una especie podría crecer de, digamos, medio metro a un metro de altura, un aumento de cien por ciento, pero a un ritmo de un milímetro por siglo, lo que me da la impresión de ser un cambio imperceptiblemente gradual.

Inventemos otros ejemplos de masbienismo, para asegurarnos de que queda clara la naturaleza del truco:

> No es que las personas sean «robots húmedos» (como dice Dilbert, y con él la mayoría de los investigadores en ciencia cognitiva); es *más bien* que tienen libre

albedrío y son moralmente responsables de sus acciones buenas y malas.

Otra vez, ¿por qué no las dos cosas? Lo que falta es un argumento que demuestre que los «robots húmedos» no pueden también ser gente con libre albedrío y moralmente responsable. Este ejemplo se aprovecha de una suposición común, pero polémica. He aquí otro:

> La religión no es el opio del pueblo, como decía Marx; es *más bien* una profunda y consoladora señal de que la humanidad reconoce la inevitabilidad de la muerte.

Una vez más, ¿por qué no puede ser tanto el opio como la señal consoladora? Creo que la idea ya ha quedado clara, y ahora para el lector será más fácil buscar masbienismos en un documento que falsas dicotomías (que nunca se anuncian como tales): basta con teclear «más bien» en el cuadro de búsqueda y ver qué sale. Recuerda: no todos los «más bien» son masbienismos; algunos son legítimos. Y algunos masbienismos no usan la expresión «más bien». He aquí uno que usa el más conciso de « _____, no _____»; lo inventé a partir de elementos que hay en la obra de varios ideólogos de la ciencia cognitiva.

> Los sistemas nerviosos deben verse como algo que activamente investiga su medio, no como simples computadoras que trabajan pasivamente a partir de los datos que obtienen de los órganos sensoriales.

¿Quién dice que las computadoras que trabajan a partir de datos que reciben no pueden investigar activamente? Este conocido contraste entre las computadoras sombríamente «pasivas» y los organismos maravillosamente «activos» nunca se ha defendido como se debe, y es uno de los más ubicuos bloqueadores de la imaginación que conozco.

Una variación del masbienismo que frecuentemente usa Gould podría llamarse *apilamiento:*

> Hablamos de la "marcha de mónada a hombre" (otra vez el lenguaje anticuado) como si la evolución hubiera seguido caminos de progreso continuos a lo largo de linajes ininterrumpidos. Nada podría estar más lejos de la realidad.[16]

¿Qué no podría estar más lejos de la realidad? Al principio parecería como si Gould dijera que no hay un linaje continuo, ininterrumpido, entre las «mónadas» (organismos unicelulares) y nosotros, pero desde luego que sí lo hay. No hay ninguna implicación de

la gran idea de Darwin más firme que ésa. ¿Entonces qué puede estar diciendo Gould? Es de suponer que debemos hacer hincapié en «caminos de *progreso*»: es (solamente) la creencia en el progreso lo que está «lejos de la realidad». Los caminos son linajes continuos e ininterrumpidos, seguro que sí, pero no son linajes de progreso (global). Esto es cierto: son linajes continuos (ininterrumpidos) de progreso (principalmente) local. Con este pasaje de Gould nos queda la sensación —a menos que seamos cautelosos— de que nos ha mostrado que algo anda verdaderamente mal en la proposición estándar de la teoría evolutiva de que hay caminos continuos (linajes ininterrumpidos) de las mónadas al hombre. Sin embargo, para usar la propia frase de Gould, «nada podría estar más lejos de la realidad».

Otro truco suyo es el *pasodoble de Gould,* un mecanismo cuya descripción publiqué hace años, y que fue así bautizado en aquel entonces por el teórico de la evolución Robert Trivers en honor de su inventor:[17]

> En la primera fase, creas al hombre de paja, y lo «refutas» (todo el mundo conoce ese truco). En segundo lugar (ésta es la genialidad), apuntas hacia las pruebas de que tú mismo diste el primer paso —las pruebas de que tus oponentes de hecho no sostienen la opinión que les atribuiste— ¡pero interpretas estas citas como concesiones a regañadientes de ellos a tu ataque![18]

En mi ensayo, una carta al director del *New York Review of Books,* publicación en la que dos meses antes Gould había criticado despiadadamente el magnífico libro de Helena Cronin *The Ant and the Peacock,*[19] presenté tres ejemplos del pasodoble de Gould. Éste el que más puede interesar aquí:

> El ejemplo más transparente es el «extrapolacionismo» que inventó Gould, descrito como una extensión lógica del «adaptacionismo de Cronin». Ésta es una doctrina de pancontinuidad y pangradualismo que el hecho de la extinción masiva refutó convenientemente —es más: trivialmente—. «Sin embargo, si las extinciones masivas son verdaderas interrupciones de la continuidad, si la lenta construcción de la adaptación en tiempos normales no se extiende al éxito predicho a través de las fronteras de la extinción masiva, entonces el extrapolacionismo falla y el adaptacionismo sucumbe». No veo por qué un adaptacionista sería tan tonto para avalar algo parecido al «extrapolacionismo» en una forma tan «pura» como para negar la posibilidad o incluso la probabilidad de que la extinción masiva tuviera un papel muy importante en la poda del árbol de la vida, como lo dice Gould. Siempre ha sido obvio que el dinosurio más perfecto sucumbirá si un cometa choca contra su tierra natal con una fuerza cientos de veces más grande que todas las bombas de

> hidrógeno que se hayan fabricado jamás. En el libro de Cronin no hay una sola palabra que apoye a Gould en su acusación de que ella cometió este error. Si Gould piensa que el papel de las extinciones masivas en la evolución es pertinente para cualquiera de los problemas fundamentales que aborda Cronin: la selección sexual y el altruismo, no dice cómo o por qué. Cuando, en el último capítulo, Cronin presenta una magnífica discusión sobre la cuestión fundamental de la teoría evolucionista en la que *no* se ha concentrado: el origen de las especies, y señala que sigue siendo un problema pendiente, Gould se abalanza y llama a esto una epifanía de último minuto, un irónico reconocimiento que hace Cronin de que su «panadaptacionismo» ha sido derrotado. ¡Qué ridiculez![20]

Idea de proyecto para un estudiante de retórica: peinar el inmenso conjunto de las publicaciones de Gould y catalogar las diferentes clases de muletas de razonamiento que explotó, empezando por el masbienismo, el apilamiento y el pasodoble.

10. El operador «no cabe duda»: un *blockeo* mental

Cuando lees o echas un vistazo a ensayos argumentativos, especialmente los escritos por filósofos, he aquí un rápido truco que te puede ahorrar mucho tiempo y esfuerzo, sobre todo en esta época en que es tan fácil hacer búsquedas con la computadora: revisa las veces que aparece en el documento la expresión «no cabe duda», y mira cada una. No siempre, ni siquiera la mayoría de las veces, pero a menudo esas tres palabras son tan prácticas como una luz intermitente para ayudarte a encontrar un punto débil en el argumento; son como una etiqueta que te advierte que probablemente ahí hay una **muleta de razonamiento**. ¿Por qué? Porque marca la mismísima orilla de aquello de lo que el autor realmente no duda y espera que tampoco los lectores duden (si el autor realmente no tuviera duda de que todos los lectores estarían de acuerdo, no valdría la pena mencionarlo). Como está a la orilla, el autor ha tenido que recurrir a su propio criterio para decidir si intenta o no demostrar el asunto en cuestión, u ofrecer pruebas a su favor, y —como la vida es corta— ha optado por una lisa aseveración, previendo, suponemos que con fundamento, que los lectores estarán de acuerdo. Es el sitio ideal para encontrar «perogrulladas» que ni se han examinado bien ¡ni son ciertas!

La primera vez que observé este útil papel del «no cabe duda» fue mientras hacía comentarios a un ensayo de Ned Block, que incluía varios ejemplos de primera calidad en contra de mi teoría de la conciencia. He aquí uno de ellos, al que le vienen muy bien las cursivas del propio Block para subrayar tipográficamente su obviedad:[21]

> *No cabe duda de que no es sino un hecho biológico acerca de la gente —no una construcción cultural— el que algunas representaciones cerebrales perseveran lo suficiente para afectar la memoria, dominar la conducta, etcétera.*[22]

Con esto pretende rechazar —sin argumentos— mi teoría según la cual la conciencia humana es algo que, en efecto, debe aprenderse, un conjunto de microhábitos cognitivos que no es seguro que estén presentes desde el nacimiento. En otro lugar dije: «Cada vez que Block diga "No cabe duda" busca lo que podríamos llamar un *blockeo* mental».[23] Block es uno de los filósofos que más derrochan y abusan del operador «no cabe duda», pero otros confían rutinariamente en él, y cada vez que lo hacen debería sonar una pequeña alarma: «Aquí es donde ocurre la prestidigitación no deliberada: se agita la falsa premisa frente a los censores y hasta se les da un codazo para que se fijen».[24]

Hace poco decidí poner a prueba de manera un poco más sistemática mi corazonada sobre el «no cabe duda». Revisé docenas de artículos —cerca de 60— sobre filosofía de la mente en ‹philpapers.org› y busqué las veces que aparecía «no cabe duda» (en inglés,

surely). La mayoría de los artículos no usaban la expresión. En los que sí la usaban (entre una y cinco veces en la muestra que revisé), la mayor parte de los casos eran claramente inocentes; unos cuantos eran, digamos, discutibles; y había seis casos en los que la alarma sonaba fuerte y clara (para mí). Desde luego, otros podrían tener un umbral de obviedad muy diferente, y a eso se debe que no me haya tomado la molestia de tabular mis «datos» en este experimento informal. A los escépticos los aliento a que lleven a cabo sus propios sondeos y vean qué encuentran. Más adelante, en el apartado 64, se desmonta un ejemplo insigne del operador «no cabe duda».

11. Preguntas retóricas

Así como deberías estar al acecho de la expresión «no cabe duda», deberías desarrollar una sensibilidad para las preguntas retóricas en cualquier argumento o polémica. ¿Por qué? Porque, como el uso del «no cabe duda», representan la impaciencia de un autor por tomar un atajo. Una pregunta retórica está flanqueada por signos de interrogación, pero no espera que se le dé respuesta. Es decir, el autor no se toma la molestia de esperar a que le respondas, pues la respuesta es tan endiabladamente obvia que te avergonzaría decirla. En otras palabras, la mayoría de las preguntas retóricas son argumentos por reducción al absurdo plegados, demasiado obvios como para necesitar que se deletreen. He aquí un buen hábito que puedes adquirir: cada vez que veas una pregunta retórica, trata —en silencio, para tus adentros— de darle una respuesta que no sea obvia. Si encuentras una buena, sorprende a tu interlocutor respondiendo la pregunta. Recuerdo una tira de *Peanuts* publicada hace años que ilustra muy bien la táctica. Charlie Brown acaba de preguntar, retóricamente, «¿Aquí quién puede decir qué está bien y qué está mal?», y en el siguiente cuadro Lucy contesta: «Yo».

12. ¿Qué es una profulidad?

Mi difunto amigo el experto en ciencias de la computación Joseph Weizenbaum anhelaba ser filósofo, y algo tarde en su carrera intentó pasar de los tecnicismos a las honduras. En una ocasión me dijo que una tarde, después de soltar con el ceño fruncido una larga perorata sobre temas elevados en la sobremesa, su hija Miriam, que entonces era una niña, dijo: «¡Guau! Mi papá acaba de decir una profulidad». ¡Qué maravilla de acuñación espontánea![25] Decidí adoptarla y darle un uso un tanto más analítico.

Una profulidad es una proposición que *parece* tanto importante como verdadera —y profunda— pero consigue este efecto por ser ambigua. En una interpretación es manifiestamente falsa, pero causaría conmoción si fuera verdadera; en la otra interpretación es verdadera pero banal. El escucha incauto capta el atisbo de verdad de la segunda interpretación y la importancia apabullante de la primera, y piensa «¡Guau! Eso es una profulidad».

He aquí un ejemplo (sugiero al lector leerlo sentado: no es poca cosa lo que viene a continuación):

Amor sólo es una palabra.

¡Oh, guau! Alucinante, de dimensiones casi cósmicas, ¿no es así? Pues no. En una interpretación, es manifiestamente falso. No estoy seguro de qué sea amor: quizá una emoción o un apego amoroso, quizá una relación interpersonal, quizá el estado más elevado que una mente humana pueda alcanzar: pero todos sabemos que no es una palabra. ¡No puedes encontrar amor en el diccionario!

Podemos hacer salir la otra interpretación aprovechando una convención que a los filósofos nos importa enormemente: cuando *hablamos acerca de* una palabra, la ponemos entre comillas, así:

«Amor» sólo es una palabra.

Esto es cierto; «amor» es una palabra del español, pero sólo una palabra, no una oración, digamos. Empieza con «A», tiene cuatro letras y aparece en el diccionario entre «amontonar» y «amoragar», que también sólo son palabras. «Hamburguesa» es sólo una palabra. «Amor» es sólo una palabra.

Pero esto no es justo, dirás. Quien haya dicho que amor sólo es una palabra seguramente quería decir algo más. Sin duda, pero no lo dijo. Quizá quería decir que «amor» es una palabra que engaña a la gente al hacerle pensar que es el término para nombrar algo maravilloso que en realidad no existe, como «unicornio», o quizá quería

decir que la palabra es tan vaga que nadie podría nunca saber si se refiere a alguna cosa, relación o acontecimiento particular. Lo cierto es que ninguna de estas afirmaciones es muy convincente. «Amor» puede ser una palabra problemática, difícil de definir, y el amor puede ser un estado del que es difícil estar seguro, pero estas afirmaciones son obvias, no especialmente informativas o profundas.

No todas las profulidades son tan fáciles de analizar. Recientemente Richard Dawkins me llamó la atención sobre una espléndida, de la autoría de Rowan Williams, el arzobispo de Canterbury, quien describió así su fe:

> Una silenciosa espera de la verdad, respirando sentado en presencia del signo de interrogación.

Le dejo al lector como ejercicio el análisis de esta profulidad.

Resumen

Una herramienta bien empuñada se vuelve tan parte de uno como las manos y los pies, sobre todo cuando hablamos de herramientas para pensar. Equipado con estas herramientas de pensamiento sencillas y para todo uso, puedes abordar las difíciles exploraciones venideras con los sentidos más despiertos: puedes ver una abertura, oír una señal de aviso, oler gato encerrado o percibir un traspiés que bien podrías haber pasado por alto sin su ayuda. También tienes algunas máximas para tener en cuenta —las reglas de Rapoport y la ley de Sturgeon, por ejemplo— que pueden decirte consejos al oído como Pepe Grillo y recordarte que debes controlar tu agresividad cuando entras audazmente al ruedo blandiendo tus armas. Sí, las herramientas de pensamiento también son armas, y la imaginería de combate es apropiada. Según parece, la competitividad es un producto secundario de la audacia y la ambición intelectuales que se necesitan para enfrentar los problemas más duros. Hemos visto que al fragor de la batalla hasta los grandes pensadores pueden recurrir a artimañas en su deseo de hacerte ver las cosas a su manera, y la crítica constructiva rápidamente puede fundirse con la burla cuando surge una oportunidad de lanzar una agudeza.

Los problemas que abordaremos son todos ellos temas candentes: el significado, la evolución, la conciencia y especialmente el libre albedrío. Sentirás cómo te van invadiendo el terror o la repugnancia mientras te acercas a las perspectivas, y ten la seguridad de que no estás solo: hasta los expertos más encumbrados son susceptibles de tener pensamiento desiderativo y pueden estar cegados a una verdad por una convicción que se apoya más en un apego emocional que en la razón. A la gente realmente le importa si tiene libre albedrío o no, cómo su mente puede residir en su cuerpo, o cómo puede haber (o incluso si es que hay) significado en un mundo compuesto de nada más que átomos y moléculas, fotones y bosones de Higgs. A la gente *debería* importarle. A fin de cuentas, ¿qué podría ser más importante que las preguntas de qué somos y qué deberíamos hacer al respecto? Así que mira por dónde caminas. Te toparás con terrenos resbaladizos, y los mapas no te lo van a advertir.

[1] Pasaje tomado de «*Surely You're Joking , Mr. Feynman!*» *Adventures of a Curious Character [¿Está usted de broma, señor Feynman? Aventuras de un curioso personaje tal como fueron referidas a Ralph Leighton]*. Copyright © 1985 Richard P. Feynman y Ralph Leighton. Citado con permiso de W. W. Norton & Company, Inc.

[2] Esto no les da a los navegantes su posición real, un punto en el globo, pero sí les da una *línea:* están en algún lugar de esa *línea de posición.* Espera unas cuantas horas hasta que el sol se haya movido un buen tramo. Ahora escoge un punto en tu línea de posición, cualquiera, y calcula a qué altura estaría ahora el sol si ese punto fuera exactamente la elección correcta. Haz la observación, compara los resultados, corrige y obtén otra línea de posición. Donde cruce tu primer ángulo de posición estará el punto en el que te encuentras. En esas horas no sólo habrá cambiado la altura del sol sino también su rumbo de aguja, así que las líneas se cruzarán en un muy buen ángulo. En la práctica, durante esas pocas horas normalmente te estás moviendo, así que adelanta tu primera línea de posición en la dirección en que te estás moviendo calculando tu velocidad y trazando una línea de posición paralela a la original. En la vida real las cosas pueden torcerse un poco, así que trata de obtener tres líneas de posición diferentes. Si todas se intersecan exactamente en el mismo punto, o eres increíblemente bueno o increíblemente suertudo, pero lo más común es que formen un pequeño triángulo, al que se le dice *sombrero de tres picos.* Imagínate en medio del sombrero, y ésa es tu nueva posición calculada.

[3] Ése es el ideal, pero, siendo como es la naturaleza humana, no siempre estamos a la altura. Uno de los problemas admitidos pero no resueltos con la práctica científica habitual es que los resultados negativos —experimentos que no revelaron lo que estaban hechos para revelar— no se publican con suficiente frecuencia. Es conocido cómo Feynman exploró y lamentó este fallo en el sistema durante su Cargo Cult Lecture, un discurso que dio en una ceremonia de graduación en Caltech en 1974, reproducido en R. Feynman, «*Surely You're Joking , Mr. Feynman!», op. cit.*

[4] Mi otro recuerdo indeleble de esa conferencia fue la zambullida de Popper en el Gran Canal. Se resbaló al salir de la lancha en el cobertizo de la isla de San Giorgio y se cayó al canal con los pies por delante; se sumergió hasta las rodillas antes de que dos diestros lancheros lo sacaran y lo pusieran en el embarcadero. Los anfitriones estaban mortificados y listos para volver a toda prisa al hotel para que el nonagenario sir Karl pudiera ponerse unos pantalones secos, pero los que traía puestos eran el único par que había llevado, ¡y estaba agendado para abrir la conferencia en menos de media hora! El ingenio italiano tomó el control del asunto y cinco minutos después tuve una visión gozosa e inolvidable: sir Karl, sentado majestuosamente en una pequeña silla exactamente a la mitad de una habitación abovedada con piso de mármol (diseñada por Palladio), rodeado de al menos media docena de muchachas en minifalda y de rodillas, entregadas a los pantalones del filósofo con ayuda de sus secadoras de pelo. Las extensiones se alargaban como rayos hacia las paredes y convertían el cuadro en una especie de margarita humana multicolor, con sir Karl al centro, serio pero sin inmutarse. Quince minutos después estaba seco y golpeando el podio con el puño para añadir énfasis a su visión dualista.

[5] Las leyes navales son notablemente complicadas; están llenas de trampas y de cláusulas de evasión de las que sólo un experto, un abogado naval, puede llevar la cuenta, así que en inglés se le dice *sea lawyer* («abogado naval») al que usa tecnicismos para evadir su responsabilidad o echarles la culpa a los demás.

[6] El torneo de Axelrod abrió el floreciente campo de la investigación teórica sobre la evolución del altruismo. Véanse R. Axelrod y W. Hamilton, «The Evolution of Cooperation», *Science,* 211: 1390-1396, 1981, y R. Axelrod, *The Evolution of Cooperation,* Basic Books, Nueva York, 1984. Hice un repaso introductorio de esto en *Darwin's Dangerous Idea: Evolution and the Meanings of Life,* Simon & Schuster, Nueva York, 1995, pp. 479-480, y más recientemente ha habido un estallido de variaciones, tanto simulaciones como experimentos, en laboratorios de todo el mundo. La maravillosamente simple aplicación de Rapoport de la idea «No te pego si tú no me pegas» es la semilla de la que nacieron todos los estudios y modelos posteriores.

[7] Esta formulación de las reglas de Rapoport es mía, hecha a partir de mi recuerdo de la correspondencia que hace muchos años mantuve con él, y que ahora por lo visto se ha perdido. Samuel Ruth recientemente me hizo notar que la fuente original de las reglas de Rapoport está en su libro *Fights, Games, and Debates,* University of Michigan Press, Ann Arbor, 1960, y en su artículo «Three Modes of Conflict», *Management Science,* 7 (3): 210-

218, 1961, que expresa la regla 1 y la atribuye a Carl Rogers, y variaciones de las demás. Mi versión es un poco más portátil y versátil.

[8] D. C. Dennett, *Freedom Evolves,* Viking Penguin, Nueva York, 2003.

[9] R. Kane, *The Significance of Free Will,* Oxford University Press, Oxford, 1996.

[10] C. L. Morgan, *An Introduction to Comparative Psychology,* W. Scott, Londres, 1894, p. 128.

[11] S. C. Meyer, *Signature in the Cell: dna and the Evidence for Intelligent Design,* HarperOne, Nueva York, 2009.

[12] T. Nagel, «Recommendation for Book of the Year», *The Times Literary Supplement,* 27, 2009.

[13] T. Nagel, «Letter to the Editor», *The Times Literary Supplement,* 1º de enero de 2010.

[14] D. Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid,* Basic Books, Nueva York, 1979, y *Metamagical Themas: Questing for the Essence of Mind and Pattern,* Basic Books, Nueva York, 1985.

[15] S. J. Gould, «Life in a Punctuation», *Natural History,* 101 (10): 10-21, 1992, p. 12.

[16] S. J. Gould, «Tires to Sandals», *Natural History,* 98 (4): 8-15, 1989, p. 14.

[17] Correspondencia personal (1993).

[18] D. C. Dennett, «Confusion over Evolution: An Exchange», *The New York Review of Books,* 40 (1-2): 43-44, 1993, p. 43.

[19] H. Cronin, *The Ant and the Peacock,* Cambridge University Press, Cambridge, 1991.

[20] D. C. Dennett, «Confusion over Evolution…», *op. cit.,* p. 43.

[21] Para no quedarse atrás, el filósofo Jerry Fodor ha adoptado la práctica de poner en cursivas sus oraciones con «sin duda», ¡y repetirlas! (*e. g.,* Jerry Fodor, *lot 2: The Language of Thought Revisited,* Oxford University Press, Oxford, 2008, p. 38), como para decir «¡Tomen, escépticos! ¡Tomen, escépticos!»

[22] N. Block, «What is Dennett's Theory a Theory Of?», *Philosophical Topics,* 22: 23-40, 1994, p. 27.

[23] D. C. Dennett, «Get Real», *Philosophical Topics,* 22: 505-556, 1994, p. 549.

[24] D. C. Dennett, «Heterophenomenology Reconsidered», *Phenomenology and Cognitive Science,* 6 (1-2): 247-270, 2007, p. 252.

[25] Hace poco Miriam encontró en internet mi uso del término [*deepity* en inglés] y se puso en contacto conmigo. Su propia versión de los hechos es un poco diferente, pero mantiene el mismo espíritu: «En mi familia la expresión conlleva algo de burla: es una idea que se disfraza de Verdad a fin de elevar al hablante». Gentilmente me permitió usar la versión que su padre me contó (según yo la recuerdo) y mi redefinición algo más estrecha de su genial neologismo.

III. Herramientas para pensar acerca del significado o contenido

¿Por qué comenzar con el significado? Empiezo con este tema porque está en el corazón de todos los problemas difíciles, por una sencilla razón: estos problemas no surgen hasta que nos ponemos a hablar de ellos, con los demás y también para nuestros adentros. A los tejones no les preocupa el libre albedrío, y ni siquiera a los delfines se les puede fastidiar con el problema de la conciencia, pues *plantear preguntas* no es parte de su repertorio. La curiosidad podrá haber matado al gato, pero a los seres humanos reflexivos nos lanza a matorrales de desconcierto. Ahora bien, tal vez esto sea el inconveniente del lenguaje, y estaríamos mejor —seríamos mamíferos más sanos y felices— si fuéramos tan ajenos a estos temas como nuestros parientes los grandes simios. Sin embargo, dado que tenemos lenguaje, tenemos que quedarnos con las Grandes Preguntas, y para bien o para mal, no nos dan la impresión de ser triviales.

El primer paso en cualquier exploración efectiva es tener tan claro como sea posible cuáles son nuestro punto de partida y nuestro equipo. Las palabras tienen significados. ¿Cómo es posible? Los usuarios de las palabras significamos cosas diciendo cosas. ¿Cómo es posible? ¿Cómo podemos entendernos unos a otros? Nuestros perros parecen ser capaces de «entender» (en cierto modo) unas cuantas palabras, incluso algunos cientos de palabras, pero fuera de esas proezas de la domesticación y de los sistemas de señales rudimentarios que se encuentran en la naturaleza (en los primates, en las aves… ¡en las jibias!), las palabras son lo que distingue nuestras mentes de las de otros animales. Es una diferencia asombrosa, pero, con todo, parece que los animales no humanos —los animales «superiores»— tienen mentes, y por lo tanto, de alguna manera quizá limitada, tratan con significados: los de sus estados perceptivos, de sus impulsos y deseos, e incluso de sus sueños.

En ocasiones los animales nos dan la impresión de ser muy parecidos a nosotros, como gente disfrazada de gato, de oso o de delfín. Esto pasa en todas las culturas humanas: se considera que los animales ven, saben, quieren, tratan, temen, deciden, desean, recuerdan, etc. En resumen, se considera que, como nosotros, tienen *mentes* llenas de… un no sé qué (¿ideas?, ¿creencias?, ¿representaciones mentales?) *significativo*. ¿Cómo puede haber significados en los cerebros? Una idea perennemente tentadora es que como las palabras tienen significados, a lo mejor las cosas significativas en nuestros cerebros —y en los de los animales— son como palabras, ordenadas en

oraciones mentales, que expresan creencias, etc. Pero si las palabras obtienen sus significados de las mentes que las profieren, ¿de dónde obtienen sus significados las palabras mentales? Los cerebros animales ¿almacenan tanto las palabras mentales como sus definiciones, en algo como un diccionario cerebral? Y si los animales —al menos los animales «superiores»— siempre han tenido cerebros llenos de palabras mentales, ¿por qué no hablan?[1] La idea de un lenguaje del pensamiento es de lo más problemática, pero nuestros pensamientos y creencias deben estar compuestos de algo. ¿Qué otra cosa podría ser?[2]

13. Asesinato en la Plaza de Trafalgar

He aquí nuestra primera bomba de intuición. Jacques mata a su tío de un disparo en la Plaza de Trafalgar y Sherlock lo detiene en el acto. Tom lee la noticia en *The Guardian,* y Boris en *Pravda.* Ahora bien, Jacques, Sherlock, Tom y Boris han tenido experiencias notablemente diferentes —por no hablar de sus vidas hasta ese momento y sus perspectivas para el futuro—, pero hay una cosa que comparten: todos creen que un francés ha cometido un asesinato en la Plaza de Trafalgar. No es que todos hayan *dicho* esto, ni siquiera «para sus adentros»; podemos suponer que *esa proposición* no «se le ocurrió» a ninguno de ellos, y aunque así hubiera sido, habría tenido una importancia muy diferente para Jacques, Sherlock, Tom y Boris. Sin embargo, todos creen que un francés cometió un asesinato en la Plaza de Trafalgar. Es una propiedad compartida que, en efecto, sólo desde un punto de vista muy limitado es visible: el punto de vista de la psicología vulgar.[3] A los psicólogos vulgares comunes y corrientes —todos nosotros— no nos cuesta ningún trabajo asignar a la gente esas características compartidas. Lo hacemos sin tener mucha idea de lo que yace entre oreja y oreja de aquellos a quienes atribuimos esas creencias. Podemos pensar que estos cuatro sujetos deben de tener también algo *más* en común —un no sé qué en sus cerebros que tiene una forma similar y de algún modo registra su creencia compartida—, pero si es así, estamos cayendo en una especulación discutible. *Puede* haber en efecto tales estructuras nerviosas en común —si resulta que los cuatro cerebros «formaron» de la misma manera la creencia de que un francés ha cometido un asesinato en la Plaza de Trafalgar—, pero esto no es necesario en absoluto, y de hecho es muy improbable, por razones que aquí exploraremos brevemente.

Las oraciones «Tengo hambre» y «J'ai faim» comparten una propiedad, a pesar de estar compuestas de letras de diferentes formas (o diferentes fonemas cuando se dicen en voz alta), en diferentes idiomas, y a pesar de tener diferentes estructuras gramaticales: *significan,* o son *acerca de,* la misma cosa: el hambre del hablante. Esta propiedad compartida, el *significado* (de las dos oraciones en sus respectivos idiomas) o el *contenido* (de las creencias que expresan), es un tema central de la filosofía y de la ciencia cognitiva. Este *ser acerca de algo* que presentan las oraciones, las pinturas, las creencias y (sin lugar a dudas) algunos estados mentales, por ejemplo, en la jerga filosófica se conoce como *intencionalidad,* desafortunada elección de término técnico, pues las personas ajenas a la filosofía automáticamente lo confunden con la idea corriente de hacer algo intencionalmente (como en la frase «¿Sus intenciones son honorables?»). Esta oración te ayudará a recordar las diferencias: un cigarro no es *acerca* de fumar, o acerca de cualquier otra cosa, a pesar de que la *intención* es que sea fumado; un cartel de «NO FUMAR» *sí es* acerca de fumar y por lo tanto presenta intencionalidad; la creencia de que atrás de ese árbol hay un atracador presenta

intencionalidad (es acerca de un atracador —posiblemente inexistente—), pero ciertamente no es intencional en el sentido ordinario (no «lo crees a propósito», simplemente se te ocurre); alejarse del árbol es intencional en el sentido ordinario, pero no es acerca de nada. Una estrategia que casi nunca falla: simplemente acostúmbrate a recordar la expresión «ser acerca de algo» cada vez que te topes con el término filosófico «intencionalidad». Fuera del acuerdo de que el significado y el contenido son fenómenos (o incluso un solo fenómeno: la intencionalidad) estrechamente relacionados entre sí y mutuamente dependientes, sigue habiendo muy poco consenso sobre lo que es el contenido (o el significado) y cuál es la mejor manera de asirlo. Por eso tenemos que abordar el tema con delicadeza. Es un festín de problemas, pero podemos acercarnos sigilosamente y comerlo a pequeños bocados.

Este ejemplo de los cuatro diferentes creyentes pretende demostrar cómo los cerebros podrían tener poco en común sin dejar de compartir una propiedad «intencional»: creer «la misma cosa». Jacques fue un testigo —de hecho, el autor del crimen—, y la familiaridad experiencial de Sherlock con el suceso es apenas un poco menos directa, pero Tom y Boris se enteraron de este hecho en condiciones sorprendentemente distintas. Hay una cantidad enorme e indefinida de maneras diferentes de adquirir la información de que un francés cometió un asesinato en la Plaza de Trafalgar, y una cantidad enorme e indefinida de maneras diferentes de usar esa información para favorecer intereses personales (responder preguntas en un concurso televisivo, ganar apuestas, despertar el interés de los turistas franceses en Londres, etc.). Si hay una buena razón para pensar que todas estas fuentes y consecuencias deben canalizarse a través de alguna estructura común en el cerebro, tarde o temprano la descubriremos, pero por el momento no nos precipitemos a sacar conclusiones.

Antes de dejar esta bomba de intuición deberíamos seguir el consejo de Doug Hofstadter y *girar las perillas* para ver qué están haciendo sus componentes. ¿Por qué escogí esa proposición en concreto? Porque necesitaba algo lo suficientemente llamativo y memorable para que se diera la noticia en diferentes idiomas a muchos kilómetros de la escena. No es así, para nada, como adquirimos la mayoría de las creencias, y vale la pena señalarlo. Por ejemplo, Jacques, Sherlock, Tom y Boris comparten una cantidad enorme e indefinida de creencias distintas que no surgieron de manera tan dramática: por ejemplo, las creencias de que las sillas son más grandes que los zapatos, que la sopa es líquida, que los elefantes no vuelan. Si trato de informarte que los salmones en su hábitat natural no usan aparatos auditivos, me dirás que ya lo sabías, pero ¿cuándo te enteraste? No naciste sabiéndolo, no formaba parte de ningún plan de estudios en la escuela, y es sumamente improbable que alguna vez en la vida hayas formulado en tu mente una oración para expresarlo. Entonces, si bien podría parecer obvio que Boris debe de haberse enterado del asesinato «simplemente subiendo» la oración rusa pertinente del

Pravda a su cerebro, y luego «traduciéndola», um, al cerebrés, no resulta nada obvio suponer que el cerebro de Boris realizó un trámite semejante (¿del *qué* al cerebrés?) en relación con el dato de los salmones.

Aquí hay otra perilla: supón que Fido, un perro, y Clyde, una paloma, también son testigos del asesinato; pueden llevarse consigo *algo* del suceso, un ajuste en sus cerebros que podría influir sobre la conducta futura, pero no sería el hecho de que un francés cometió un asesinato en la Plaza de Trafalgar, aunque la información que apoya ese hecho podría estar presente en la luz y el sonido que incidieron en sus órganos sensoriales (una videograbación del suceso podría ofrecer pruebas en un proceso legal, por ejemplo, pero a Fido y a Clyde éstas se les escaparían). Así, esta bomba de intuición corre el riesgo de llevar un sesgo seriamente antropocéntrico a nuestra exploración del significado. Las palabras y las oraciones son vehículos de significado ejemplares, pero en cuanto a los animales que no las usan, la idea de que sus *cerebros* de todas maneras las usan es por lo menos rocambolesca… lo cual no la hace falsa. Si resulta ser verdadera, será un descubrimiento para quedarse con la boca abierta, pero ya hemos tenido otros de ésos.

Los fenómenos de intencionalidad nos son completamente familiares —tan notables en nuestra vida cotidiana como la comida, los muebles y la ropa— y al mismo tiempo son sistemáticamente escurridizos para una perspectiva científica. Tú y yo rara vez tenemos alguna dificultad para distinguir una felicitación de cumpleaños de una amenaza de muerte o de una promesa, pero piensa en la tarea de ingeniería que sería hacer un detector de amenazas de muerte confiable. ¿Qué tienen en común todas las amenazas de muerte? Al parecer, sólo su significado. Y el significado *no* es como la radiactividad o la acidez, una propiedad que un detector bien calibrado inmediatamente pueda distinguir. Lo más cerca que hemos estado hasta ahora de crear un detector de significado de uso general es Watson, de IBM, que es mucho mejor para clasificar significados que cualquier sistema de inteligencia artificial anterior, pero hay que observar que no es para nada sencillo, y (probablemente) cometería errores a la hora de identificar algunos candidatos a amenaza de muerte que en cambio un muchacho captaría de inmediato. Hasta los niños pequeños reconocen que cuando un niño le grita riendo a otro «No respondo: ¡si vuelves a hacer eso te mato!», en realidad no es una amenaza de muerte. Tan sólo el tamaño y la sofisticación de Watson son al menos indicadores indirectos de cuán escurridiza es la propiedad del significado con la que, por otra parte, estamos tan familiarizados.

14. Un hermano mayor que vive en Cleveland

Con todo, el significado no es una propiedad completamente misteriosa. De una u otra manera, las estructuras de nuestros cerebros en cierto modo «almacenan» nuestras creencias. Cuando aprendes que los pudúes son mamíferos, *algo* tiene que haber cambiado en tu cerebro; algo tiene que quedar relativamente inculcado de una manera en que antes de que lo aprendieras no lo estaba, y sea lo que sea, debe ser *acerca de algo:* lo suficiente para, de un modo u otro, explicar tu recién descubierta capacidad de identificar a los pudúes como animales más emparentados con los búfalos que con las barracudas. Así que, en efecto, es tentador imaginar que las creencias están «almacenadas en el cerebro» más o menos como los archivos de datos están almacenados en los discos duros, en algún sistema de codificación, que puede ser tan diferente de un individuo a otro como las huellas digitales. Las creencias de Jacques estarían escritas en su cerebro en jacquiano, y las de Sherlock en sherlockiano. Con todo, esta atractiva idea tiene sus problemas.

Supongamos que hemos llegado a la edad de oro de la neurocriptografía y se ha vuelto posible que un «microneurocirujano cognitivo» haga un poco de hojalatería e *inserte* una creencia en el cerebro de una persona, escribiendo la correspondiente proposición en sus neuronas y usando el lenguaje cerebral local, desde luego (si podemos aprender a *leer* la escritura cerebral, es de suponer que podemos *escribir* escritura cerebral, si nuestras herramientas son lo suficientemente delicadas). Supongamos que vamos a insertar en el cerebro de Tom la siguiente creencia falsa: *Tengo un hermano mayor que vive en Cleveland.* Supongamos que el microneurocirujano cognitivo puede hacer el recableado necesario, tan a fondo y tan delicadamente como queramos. Este recableado puede dañar o no dañar la racionalidad básica de Tom; consideremos los dos resultados. Tom está sentado en un bar y un amigo le pregunta: «¿Tienes hermanos o hermanas?» Tom dice: «Sí, tengo un hermano mayor que yo y vive en Cleveland». «¿Cómo se llama?» ¿Qué ocurrirá ahora? Tom puede responder: «Se llama… ¿Cómo se llama quién? Dios mío, ¿qué estaba diciendo? ¡No tengo ningún hermano mayor! ¡Por un instante me pareció tener un hermano mayor que vive en Cleveland!» Otra posibilidad es que diga: «No sé cómo se llama», y si le insisten negará cualquier conocimiento acerca de este hermano y dirá cosas como «Soy hijo único y tengo un hermano mayor que vive en Cleveland». En ninguno de los casos consiguió nuestro microneurocirujano cognitivo, con su recableado, implantar una nueva creencia. En el primer caso, la racionalidad intacta de Tom borra a la intrusa (solitaria, carente de apoyo) en cuanto surge. Una disposición evanescente a decir «Tengo un hermano mayor que yo y vive en Cleveland» no es realmente una creencia: es algo de naturaleza más parecida a la de un tic, como una manifestación del síndrome de Tourette. Y si persiste esta patología en el pobre Tom, como en la segunda alternativa, su franca irracionalidad en el tema de los hermanos

mayores lo descalifica como creyente. Nadie que entienda que no se puede ser hijo único y tener un hermano mayor que vive en Cleveland realmente entiende la oración que Tom pronunció, y lo que realmente no se entiende no se puede creer, aunque sí se pueda repetir como loro.

Este ejemplo de ciencia ficción pone de relieve la suposición tácita de competencia mental que subyace en todas las atribuciones de creencia; a menos que se tenga un repertorio indefinidamente extensible de maneras de *usar* tu creencia candidata (si creencia es lo que es) en diferentes contextos, no es una creencia en ningún sentido remotamente reconocible. Si el cirujano ha hecho la tarea con delicadeza y ha preservado la competencia del cerebro, ese cerebro deshará el trabajo en cuanto surja el tema; de lo contrario, patológicamente, rodeará la obra con capas de preciosos recuerdos fabricados («Se llama Sebastián, y es un acróbata de circo que vive en un globo»). Tal fabricación de recuerdos no es desconocida; quienes sufren del síndrome de Korsakov (la amnesia que frecuentemente aqueja a los alcohólicos) pueden ser asombrosamente convincentes cuando relatan sus pasados «recordados» pero sin pizca de verdad. Esta misma elaboración, sin embargo, es prueba clara de que la persona no *nada más* tiene una «proposición» aislada almacenada en el cerebro; incluso una creencia delirante *requiere* el apoyo de una gran cantidad de creencias no delirantes y la capacidad de reconocer las implicaciones de todo esto. Si no cree que su hermano mayor también es un hombre, respira, vive al oeste de Boston, al norte de Panamá, etc., etc., entonces decir que la hazaña del cirujano fue insertar una *creencia* sería más que engañoso.

Lo que esta bomba de intuición muestra es que nadie puede tener sólo una creencia (no puedes creer que un perro tiene cuatro patas sin creer que las patas son extremidades y que cuatro es mayor que tres, etc.).[4] También muestra otras cosas, pero no me detendré en enumerarlas. Tampoco intentaré decir cómo podría usarse para otros fines una variación de esta herramienta de pensamiento específica, aunque el lector está invitado a girar las perillas por su cuenta y ver qué se le ocurre. Quiero haber expuesto un surtido variado de este tipo de herramientas de pensamiento antes de que reflexionemos más sobre sus características.

15. «Mi papá es doctor»

Le preguntan a una niña a qué se dedica su padre y responde «Mi papá es doctor». ¿Cree lo que dice? En un sentido, por supuesto, pero ¿qué tendría que saber para *realmente* creerlo? (¿Y si dijera «Mi papá es arbitrajista» o «Mi papá es actuario?») Supongamos que tenemos la sospecha de que la niña habla sin entender, y decidimos ponerla a prueba. ¿Debe poder producir paráfrasis o explayarse en su afirmación diciendo que su padre cura a la gente enferma? ¿Es suficiente con que sepa que el hecho de que su papá sea un doctor descarta que sea un carnicero, un pastelero, un fabricante de velas? ¿Sabe qué es un doctor si le falta el concepto de un médico farsante, de un curandero o de alguien que ejerce la medicina sin autorización? Es más, ¿cuánto necesita entender para saber que su papá es su padre? (¿Su padre adoptivo? ¿Su padre «biológico»?) Obviamente su entendimiento de lo que es ser un doctor, así como de lo que es ser un padre, crecerá con los años, y por ende su entendimiento de su propia oración «Mi papá es doctor» también crecerá. ¿Podemos especificar —de un modo no arbitrario— cuánto debe saber a fin de entender esta proposición «completamente»? Si la comprensión viene en grados, como muestra este ejemplo, entonces la creencia, que depende del entendimiento, también debe venir en grados, incluso para proposiciones tan mundanas como ésta. Ella maomeno cree que su padre es un doctor, lo cual no quiere decir que tenga reservas o dudas, sino que no alcanza el nivel de entendimiento que es una condición importante para cualquier concepto útil de creencia.

16. Imagen manifiesta e imagen científica

Es momento de levantar algún montaje antes de seguir adelante con nuestro intento de comprender qué son los significados. He aquí una herramienta de pensamiento que ofrece una perspectiva valiosa de tantos temas que debería formar parte del juego de herramientas de todo el mundo, pero hasta ahora no se ha difundido muy lejos de su hogar en la filosofía. El filósofo Wilfrid Sellars la concibió en 1962 para facilitarnos el cavilar sobre qué nos enseña la ciencia acerca del mundo en que vivimos. La imagen *manifiesta* del mundo tal como se nos aparece en la vida cotidiana, llena de objetos sólidos, colores y olores y sabores, voces y sombras, plantas y animales, y gente y todas sus cosas: no sólo mesas y sillas, puentes e iglesias, dólares y contratos, sino también cosas intangibles como canciones, poemas, oportunidades y libre albedrío. Piensa en todas las preguntas desconcertantes que surgen cuando tratamos de poner en fila todas esas cosas junto con las de la imagen *científica:* moléculas, átomos, electrones, quarks y demás ralea. ¿Hay algo que *realmente* sea sólido? A principios del siglo XX, el físico sir Arthur Eddington escribió acerca de las «dos mesas», la sólida de la experiencia cotidiana y la que está compuesta de átomos, muy separados entre sí por espacio vacío en la mayor parte: algo más parecido a una galaxia que a un pedazo de madera. Algunas personas decían que lo que mostraba la ciencia era que *nada era realmente sólido,* que la solidez era una ilusión, pero Eddington era suficientemente sensato para no ir tan lejos. Algunas personas han dicho que el color es una ilusión. ¿Lo es? La radiación electromagnética en la estrecha gama a la que responde la visión humana (la gama entre el infrarrojo y el ultravioleta) no está hecha de cositas de colores, y los átomos no tienen color, ni siquiera los átomos del oro. Aun así, el color no es una ilusión en el sentido que nos importa: nadie piensa que Sony está mintiendo cuando dice que sus televisiones a color realmente muestran el mundo de color, o que debería demandarse a Sherwin-Williams por fraude por vendernos tantos colores diferentes en forma de pintura. ¿Y qué decir de los dólares? Hoy en día, la inmensa mayoría de los dólares no están hechos de plata o siquiera de papel. Son *virtuales,* hechos de información, no de material, tal como los poemas y las promesas. ¿Significa eso que sean una ilusión? No, pero no los busques entre las moléculas.

Es famosa esta afirmación de Sellars: «El objeto de la filosofía, formulado de manera abstracta, es entender cómo las cosas en el más amplio sentido del término encajan en el más amplio sentido del término».[5] Es la mejor definición de la filosofía que he encontrado. La tarea de resolver cómo cuadrar todas las *cosas* conocidas de nuestra imagen manifiesta con todas las *cosas* relativamente desconocidas de la imagen científica no es un trabajo para el que los científicos estén especialmente preparados. Por favor dígame, doctor en física, ¿exactamente qué es un *color*? ¿Existen los colores de acuerdo con su teoría? Doctor en química, ¿puede usted darnos la fórmula química de un *trato*?

No cabe duda (¡ring!) de que hay tratos. ¿De qué están hechos? Mmmh. A lo mejor no hay tratos, no en realidad. ¿Pero entonces cuál es la diferencia —la diferencia *química*— entre algo que es un trato y algo que sólo parece ser un trato? Podríamos seguir en esta misma vena, examinando gran cantidad de enigmas que sólo los filósofos han hecho esfuerzos serios por resolver, pero mejor vayamos un paso atrás, como nos propone Sellars, y fijémonos en el hecho de que hay estas dos perspectivas del mundo notablemente diferentes. ¿Por qué hay dos? ¿O habrá muchas? Tratemos de responder esta pregunta comenzando con la imagen científica y viendo si desde esa posición estratégica divisamos el surgimiento de la imagen manifiesta.

Todo organismo, ya sea una bacteria o un miembro de la especie *Homo sapiens,* tiene en el mundo un conjunto de cosas que le importan y que por lo tanto necesita discriminar y prever lo mejor que pueda. Los filósofos llaman a la lista de cosas que se considera que existen una *ontología* (de la palabra griega para «cosa», ¡oh sorpresa!). Todo organismo, por tanto, tiene una ontología. Otro nombre para ello es la *Umwelt* del organismo[6] (*Umwelt* es la palabra alemana para «medio ambiente», ¡oh sorpresa!). La *Umwelt* de un animal consiste, en primer lugar, de *proporcionamientos [affordances]:*[7] cosas que comer, o con las cuales aparearse, o a las cuales rehuir, brechas que atravesar o por las cuales asomarse, agujeros para esconderse, cosas en las que pararse, etc. La *Umwelt* de un organismo es en un sentido un medio ambiente *interno,* una ontología «subjetiva» e incluso «narcisista», compuesta sólo de las cosas que más le importan a *él,* pero su *Umwelt* no es necesariamente interna o subjetiva en el sentido de ser consciente. La *Umwelt* en realidad es un concepto ingenieril; piensa en la ontología de un elevador manejado por computadora, el conjunto de todas las cosas a las que tiene que dar seguimiento para hacer su trabajo.[8] Uno de los estudios de Von Uexküll fue sobre la *Umwelt* de una garrapata. Podemos suponer que la *Umwelt* de una estrella de mar, un gusano o una margarita es más parecida a la ontología de un elevador que a *nuestra Umwelt,* que es, de hecho, nuestra imagen manifiesta.

Nuestra imagen manifiesta, a diferencia de la ontología o *Umwelt* de la margarita, realmente es manifiesta, realmente es subjetiva en un sentido fuerte. Es el mundo en el que vivimos, el mundo según nosotros.[9] Sin embargo, igual que la ontología de la margarita, mucho de nuestra imagen manifiesta ha sido formado por la selección natural a lo largo de millones de años y es parte de nuestra herencia genética. Uno de mis ejemplos favoritos de cuán diferente puede ser una *Umwelt* compara un oso hormiguero con un pájaro insectívoro.[10] El pájaro sigue el vuelo de insectos individuales y lidia con sus trayectorias erráticas gracias a un elevado umbral de fusión de parpadeo (de hecho ve más «cuadros por segundo» que nosotros, y en consecuencia para él una película sería como una proyección de diapositivas). El oso hormiguero sólo echa un cálculo sobre un área infestada de hormigas y deja que su gran lengua atrape todas las que pueda. Un

filósofo podría decir que para un oso hormiguero «hormiga» es un *término masa,* como «agua», «hielo» y «mobiliario», no un *clasificador* (el que se puede contar), como «aceitunas», «gotas» o «sillas». Cuando un oso hormiguero ve una buena mancha de *hormiga,* la sorbe con la lengua, tan ajeno a los individuos como nosotros a las moléculas de glucosa que sorbemos cuando comemos un dulce.

La mayor parte de nuestra imagen manifiesta no viene con la herencia genética; se nos inculca de alguna manera en nuestra experiencia en la primera infancia. Las palabras son para nosotros una muy importante categoría de cosas, y son el medio a través del cual se transmite gran parte de nuestra imagen manifiesta, pero la capacidad de categorizar algunos sucesos del mundo como palabras, y nuestro deseo de hablar, bien puede ser un talento que nos venga al menos parcialmente por herencia genética: como la capacidad del pájaro para distinguir insectos voladores individuales, o el deseo de una avispa por cavar un nido. Incluso sin una gramática que teja unas con otras para formar oraciones, las palabras como simples rótulos pueden ayudar a generar importantes categorías bien enfocadas en nuestra imagen manifiesta: mamá, perrito, galleta. ¿Podrías armar un concepto claro de lo que es un trato, un error o una promesa —ya ni digas un jonrón o un soltero— sin ayuda de las palabras que acabo de usar para mencionarlos? Nuestra consideración de la lista de las herramientas de pensamiento favoritas de Doug Hofstadter ya nos ha mostrado cómo los términos pueden estructurar y sazonar nuestras mentes, enriqueciendo nuestras imágenes manifiestas personales con *cosas* —balas perdidas, dientes afuera y retroalimentación— que de otro modo son invisibles.

17. Psicología vulgar

Probablemente el motivo más importante de nuestra imagen manifiesta, porque ancla tantas otras categorías que nos interesan, es lo que llamo *psicología vulgar*. En 1981 acuñé la expresión con su significado actual, pero según parece tuvo una encarnación anterior en los escritos de Wilhelm Wundt, Sigmund Freud y otros (*Volkpsychologie* en alemán; *folk psychology* en inglés), donde significaba algo relacionado con el carácter nacional (el *Geist* del *Volk* alemán —mejor no quieres saberlo—). Este antecedente se me escapó en su momento, como a muchos otros que adoptaron el término. Lo propuse para nombrar el *talento* que todos tenemos para interpretar a la gente que nos rodea —y a los animales, robots y hasta los humildes termostatos— como *agentes* que tienen información sobre el mundo en el que actúan *(creencias)* y sobre las metas *(deseos)* que se esfuerzan por alcanzar, y que dadas sus creencias y deseos eligen el curso de acción más *razonable*.

A algunos investigadores les gusta llamar «teoría de la mente» a la psicología vulgar, pero a mí me parece engañoso, pues tiende a prejuzgar la cuestión de cómo nos las arreglamos para tener dicho talento al sugerir que tenemos una teoría y la aplicamos. Con el mismo espíritu, tendríamos que decir que alguien tiene una *teoría de la bicicleta* si sabe andar en bicicleta, y que la *teoría de la nutrición* de la gente explica su capacidad de evitar morirse de hambre y abstenerse de comer arena. No me parece que sea una manera útil de pensar acerca de estas competencias. Como todo el mundo está de acuerdo en que tenemos el talento interpretativo, y no todo el mundo está de acuerdo en cómo nos las arreglamos para ser tan competentes, creo que es mejor dejar fuera la palabra «teoría» y por el momento usar un término un poco más neutral. También la psicología académica o científica se ocupa de explicar y predecir las mentes de otros, y ella sí que tiene teorías: conductismo, cognitivismo, modelos neuroinformáticos, psicología Gestalt, entre muchas otras. La psicología vulgar es un talento en el que destacamos sin una educación formal. La *física vulgar,* de manera paralela, es el *talento* que tenemos para esperar que los líquidos fluyan, que las cosas sin apoyo se caigan, que las sustancias calientes nos quemen, que el agua nos quite la sed y que las piedras rodantes no acumulen musgo. Otra pregunta interesante es cómo se las arreglan nuestros cerebros para generar *estas* expectativas casi siempre correctas con tan poco esfuerzo, aunque nunca hayamos tomado un curso de física.

La psicología vulgar es «lo que todo el mundo sabe» de su mente y de las mentes de otros: la gente puede sentir dolor o tener hambre o sed y reconocer la diferencia; puede recordar sucesos de su pasado, prever un montón de cosas, ver qué hay frente a sus ojos abiertos, oír lo que se dice lo suficientemente cerca, engañar y ser engañada, saber dónde está, reconocer a otros, etc. La confianza con que hacemos estas suposiciones es impresionante, dado lo poco que sabemos acerca de lo que realmente pasa adentro de las

cabezas de estas personas (por no hablar de otros animales). Estamos tan seguros de todo esto que se necesita un enorme distanciamiento para siquiera darnos cuenta de que lo estamos haciendo.

Los artistas y los filósofos están de acuerdo en una cosa: una de las tareas que se han asignado a sí mismos es «hacer que lo conocido resulte extraño».[11] Algunas de las grandes genialidades creativas nos hacen atravesar la corteza de la familiaridad excesiva, **pensar afuera de la caja** sobre la distinta perspectiva desde la que podemos ver cosas obvias y ordinarias con una nueva mirada. Los científicos están completamente de acuerdo. El momento mítico de Newton fue cuando se planteó a sí mismo la extraña pregunta de por qué la manzana *cayó* del árbol («¿Y por qué no *había* de caer?», se pregunta el no genio común y corriente, «¡es *pesada*!»: como si esto fuera una explicación satisfactoria). Si no eres ciego y tienes amigos ciegos, probablemente confirmarás mi corazonada de que sin importar cuánto tiempo pases en su compañía, *todavía* te descubres de pronto a ti mismo usando las manos para señalar cosas o para trazar formas explicativas frente a sus ojos que no ven, a fin de hacerles creer y entender lo que tú crees y entiendes. La expectativa «por defecto» cuando estás en presencia de otro ser humano despierto es que ambos pueden ver, oír y oler las mismas cosas. Vas manejando por la carretera a 100 kilómetros por hora, impasible frente al hecho de que otro coche viene hacia ti en el carril contrario a la misma velocidad. ¿Cómo sabes que no habrá un choque terrible? Maquinalmente asumes que el conductor (a quien ni siquiera puedes ver y es casi seguro que no conozcas) *quiere* seguir vivo y *sabe* que la mejor manera de hacerlo es quedarse en el lado derecho del camino. Observa que todo lo que se necesitaría para ponerte en un estado mental más ansioso sería una noticia en el radio que informara al mundo que ese día en tu ruta estaba poniéndose a prueba un nuevo sistema de conducción de carros robótico. Sí, está *diseñado* para ser seguro, y después de todo Google lo creó, pero tu confianza en la racionalidad y el conocimiento del chofer humano en dirección contraria promedio es mayor que la que puedas tener en el cacareado robot (excepto quizás los sábados ya entrada la noche).

¿Cómo es que «todo el mundo conoce» la psicología vulgar? ¿A fin de cuentas sí es una especie de teoría de la mente que se aprende de niños? ¿Algo de ella es innato? O si la aprendemos toda, ¿cómo y cuándo lo hacemos? En los últimos años ha habido una oleada de investigaciones alrededor de estas preguntas. Mi respuesta (todavía, después de todos estos años de investigación) es que no es tanto una teoría como una práctica, una manera de investigar el mundo que nos sale tan natural que debe de tener alguna base genética en nuestros cerebros. Una parte de esto sí tuvimos que «mamarla», y alguien que de alguna manera hubiera estado privado de todo contacto con seres humanos durante su crecimiento tal vez sería asombrosamente inepto en psicología vulgar (aparte de tener otras graves incapacidades), pero tenemos un muy fuerte impulso de interpretar

las cosas que se mueven de maneras irregulares (a diferencia de un péndulo o una pelota rodando hacia abajo en una colina) como si fueran agentes. Una demostración de esto que gusta mucho en los cursos introductorios de psicología es una breve animación (con la técnica cuadro por cuadro conocida como *stop motion*) hecha por Fritz Heider y Mary-Ann Simmel en 1944. Muestra dos triángulos y un círculo moviéndose de un lado a otro, dentro y fuera de una caja. Las formas geométricas no podrían ser menos parecidas a la gente (o a los animales), pero es casi imposible resistirse a ver sus interacciones como algo que tiene un sentido, movidas por deseo, miedo, valentía y enojo (véase la figura 1). Este antiguo demo puede encontrarse en muchos sitios web; basta con poner «Heider y Simmel» en un buscador.

Nacemos con un «dispositivo de detección de agentes»,[12] y se activa a la menor provocación. Cuando falla, lo que a menudo ocurre en circunstancias estresantes, tendemos a ver fantasmas, duendes, diablillos, hadas, gnomos, genios y otros seres por el estilo ahí donde todo lo que realmente hay son ramas agitadas por el viento, paredes de piedra inclinadas o puertas chirriantes.[13] Desde muy niños, involuntariamente y sin

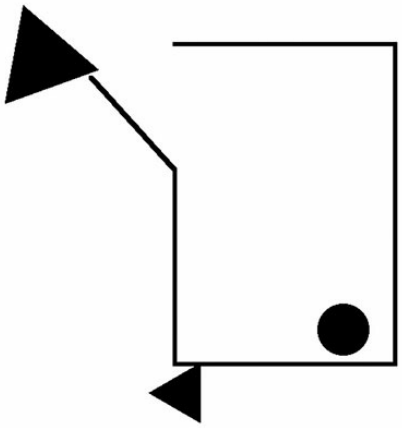


FIGURA 1

esfuerzo, vemos a los demás como agentes, y no sólo alegres, enojados, desconcertados o temerosos, sino como alguien a quien *ya le contaron el secreto,* que *se pregunta hacia dónde dar la vuelta* o incluso alguien que *no quiere aceptar el trato que le proponen.* No es neurocirugía ni se necesita ser un genio; es fácil. El poder y la facilidad de la psicología vulgar se deben, como he argumentado, a las suposiciones simplificadoras que la hacen posible. Es *como* un modelo idealizado en la ciencia: máximamente abstracto y reducido a lo esencial. Lo llamo la *postura intencional.*[14]

18. La postura intencional

Hasta aquí todo bien. Casi todos somos buenos con la psicología vulgar.[15] Tenemos la habilidad de pensar que los demás tienen mentes (y también nosotros mismos), una práctica que la mayor parte del tiempo requiere tan poco esfuerzo como la respiración. Nos basamos en ella sin el menor reparo y sin pensarlo dos veces, y es formidablemente confiable. ¿Por qué es tan fácil y cómo funciona? Aquí tenemos que hacer una pausa y erigir un andamiaje que nos haga mucho más fácil la vida mientras avanzamos en nuestras exploraciones.

¿Cómo funciona la psicología vulgar? Podemos vislumbrar bien la respuesta observando cómo funciona cuando la aplicamos a cosas que *no son* otras personas. Imagina que estás jugando ajedrez contra una computadora. Quieres ganar, y la única buena manera de acercarte a ese fin es tratar de prever las respuestas de la computadora a tus jugadas: «Si moviera mi alfil hacia allá, la computadora se lo comería; si mejor moviera el peón, la computadora tendría que mover su dama…» ¿Cómo sabes lo que hará la computadora? ¿Has visto dentro de ella? ¿Has estudiado su programa de ajedrez? Claro que no. No tienes que hacerlo. Haces tus confiadas predicciones a partir de las muy obvias suposiciones de que la computadora:

1. «conoce» las reglas y «sabe cómo» se juega el ajedrez;
2. «quiere» ganar, y
3. «tomará» estas posibilidades y oportunidades como lo que son, y actuará en consecuencia (es decir, racionalmente).

En otras palabras, supones que la computadora es una *buena* jugadora de ajedrez, o al menos no una jugadora estúpida y autodestructiva. La tratas, en *otras* palabras, como si fuera un ser humano con una mente. Incluso en otras palabras, cuando usas la psicología vulgar para prever y entender sus jugadas, *has adoptado la postura intencional.*

La postura intencional es la estrategia de interpretar la conducta de una entidad (persona, animal, artefacto o lo que sea) tratándola *como si* fuera un agente racional que determina su «elección» de «acción» «considerando» sus «creencias» y «deseos».[16] El entrecomillado de todos estos términos llama la atención sobre el hecho de que algunas connotaciones habituales podrían dejarse de lado a fin de explotar sus características principales: su papel en el razonamiento práctico, y por lo tanto en la predicción de la conducta de razonadores prácticos. Cualquier cosa que sea útil y ampliamente predecible desde la postura intencional es, por definición, un *sistema intencional,* y como veremos, muchas cosas fascinantes y complicadas que no tienen cerebros, ojos, oídos o manos, y que por lo tanto en realidad no tienen mentes, son, no obstante, sistemas intencionales.

Es decir, el truco elemental de la psicología vulgar tiene algunas aplicaciones adicionales fuera del mundo de las interacciones humanas (veremos aplicaciones no sólo en tecnología informática y neurociencia cognitiva, sino también en biología evolutiva y del desarrollo, por mencionar las áreas más importantes).

Propongo que simplemente pospongamos la preocupante pregunta de qué *de verdad* tiene una mente, de cuál es el dominio *adecuado* de la postura intencional. Sea cual sea la respuesta correcta a esa pregunta —si es que tiene una respuesta correcta—, esto no pondrá en peligro el hecho simple y llano de que la postura intencional *funciona* extraordinariamente bien como método de predicción en estas otras áreas, casi tan bien como en nuestras vidas cotidianas de psicólogos vulgares que tratan con otras personas. Esta jugada mía irrita y frustra a algunos filósofos que quieren llamar al orden e insisten en resolver como es debido el tema de qué *es una mente, una creencia o un deseo, antes de dar el siguiente paso. ¡Defina sus términos, señor!* No, no lo haré. Sería prematuro. Primero quiero explorar el poder y el alcance de la aplicación de este buen truco, la postura intencional. Cuando hayamos visto para qué sirve, y por qué, podremos regresar y preguntarnos si aún sentimos la necesidad de tener definiciones formales, herméticas. Mi jugada es un ejemplo de *picar* un problema difícil en vez de tratar de comérselo completo (y digerirlo) desde el principio. Muchas de las herramientas de pensamiento de las que estaré haciendo demostraciones son buenas para picar, para localizar, a grandes rasgos, algunos puntos «fijos» que nos ayuden a ver la forma general del problema. En *La libertad de acción. Un análisis de las exigencias del libre albedrío*[17] comparé mi método con el del escultor cuando bosqueja la forma en un bloque de mármol y se acerca a las superficies finales con cautela, modestamente, trabajando en aproximaciones sucesivas. Por lo visto, muchos filósofos no pueden trabajar así y tienen que fortificar (o eso piensan) las completamente fijas fronteras de sus problemas y posibles soluciones antes de poder aventurar cualquier hipótesis.

Veamos entonces de dónde viene el poder de la postura intencional comparándola con otras tácticas de previsión. Comencemos por identificar tres posturas fundamentales (que podrían tener otras subdivisiones para algunos propósitos, que sin embargo no son los nuestros en este momento): la postura física, la postura de diseño y la postura intencional.

La *postura física* es simplemente el laborioso método estándar de las ciencias físicas, en el que usamos todo lo que sabemos acerca de las leyes de la física y la constitución física de las cosas en cuestión para trazar nuestras predicciones. Cuando predigo que una piedra que se suelta de mi mano caerá al suelo, estoy usando la postura física. En general, para cosas que ni están vivas ni son artefactos, la postura física es la única estrategia disponible, si bien hay excepciones importantes, como veremos. Todas las cosas físicas, sean diseñadas o vivas, o no, están sujetas a las leyes de la física y por lo tanto se comportan de maneras que *en principio* pueden explicarse y predecirse desde la postura física. Si la cosa que suelto es un despertador o un pececito, hago la misma predicción acerca de su trayectoria descendente, sobre la misma base. Predecir las conductas más interesantes de los despertadores y los pececitos desde la postura física casi nunca es práctico.

Los despertadores, como son objetos diseñados (a diferencia de una piedra), también se prestan a un estilo de predicción más elaborado: la predicción desde la *postura de diseño*. Imagínate que categorizo un objeto novedoso como un despertador: rápidamente puedo razonar que *si* aprieto de determinada manera unos cuantos botones, *entonces* cierto tiempo después el despertador hará un ruido fuerte. No necesito entender las leyes físicas concretas que explican esta maravillosa regularidad; simplemente *supongo* que tiene un diseño particular —el diseño que llamamos un despertador— y que funcionará como debe, como fue diseñado para funcionar. Las predicciones desde la postura de diseño son más arriesgadas que las de la postura física dadas las suposiciones adicionales que tienen que tenerse en cuenta:

1. que una entidad *está* diseñada como supongo que lo está, y
2. que funcionará de acuerdo con ese diseño; es decir, que no fallará.

En ocasiones las cosas diseñadas están mal diseñadas, y a veces se rompen. Nada de lo que le pasa a una piedra, o en ella, cuenta como falla, pues para empezar no tiene ninguna función, y si se rompe en dos, el resultado son dos piedras, no una piedra rota. Cuando una cosa diseñada es realmente complicada (una motosierra, por ejemplo, en contraste con un hacha), el precio moderado que uno paga en términos de riesgo se compensa con creces por la enorme facilidad de predicción. Nadie preferiría recurrir a las

leyes fundamentales de la física para predecir el funcionamiento de la motosierra en lugar de consultar un práctico diagrama de sus partes móviles.

Una postura más arriesgada y más veloz es nuestro tema central, la *postura intencional,* una subespecie de la postura de diseño en la que la cosa diseñada se trata como un tipo de agente, con creencias y deseos, y suficiente racionalidad para hacer lo que debería dadas esas creencias y deseos. Un despertador es tan simple que este imaginativo antropomorfismo es, estrictamente hablando, innecesario para nuestro entendimiento de por qué hace lo que hace, pero adoptar la postura intencional es más útil —de hecho casi obligatorio— cuando el artefacto en cuestión es mucho más complicado que un despertador. Expliquemos lentamente y en detalle la práctica de adoptar la postura intencional, concentrándonos en nuestro ejemplo de una computadora que juega al ajedrez, para asegurarnos de que no estamos pasando por alto nada importante:

- En primer lugar, haz una lista de las jugadas legales que la computadora tiene disponibles cuando llega su turno (normalmente habrá varias decenas de candidatas).
- Ahora clasifica las jugadas legales de la mejor (la más sensata, la más racional) a la peor (la más tonta, la más contraproducente).
- Por último, haz tu predicción: la computadora hará la mejor jugada.

Bien puede ser que no estés seguro de cuál es la mejor jugada (¡la computadora puede «apreciar» la situación mejor que tú!), pero casi siempre puedes eliminar todas las jugadas candidatas excepto cuatro o cinco, lo que sigue dándote una enorme ventaja predictiva. *Podrías* mejorar esta ventaja y predecir con anticipación exactamente qué movida hará la computadora —con un enorme costo de tiempo y esfuerzo— recurriendo a la postura de diseño. Obtén el «código fuente» (véase el apartado 27) del programa y «simúlalo a mano», repasando los millones o miles de millones de pequeños pasos que la computadora dará mientras encuentra su respuesta a la jugada que estás pensando hacer. Esto definitivamente te dirá qué haría la computadora si tú hicieras tu jugada, pero se te acabaría el tiempo —de hecho, tu vida habría llegado a su fin— mucho antes de que alcanzaras la conclusión. ¡Es demasiada información! De todas maneras, usar la postura de diseño es *mucho* más fácil que retroceder todo el camino hasta la postura física y tratar de calcular el flujo de electrones que resulta de presionar las teclas de la computadora. Así, pues, la postura física es completamente impráctica cuando se intenta predecir y explicar las jugadas de la computadora, e incluso la postura de diseño implica demasiado trabajo, a menos que consigas otra computadora que lo haga por ti (y eso es trampa). La postura intencional suaviza todos esos laboriosos cálculos y recogida de información y arriesgadamente se decide por una apuesta bastante buena: la computadora

será tan «racional» como para encontrar y hacer la mejor jugada (dado lo que quiere —ganar— y lo que sabe —las posiciones y valores de todas las piezas en el tablero—). En muchas situaciones, especialmente cuando la mejor jugada que puede hacer la computadora es tan obvia que se considera una «jugada forzada» o cosa fácil, la postura intencional puede predecir su jugada con una exactitud casi perfecta sin mayor esfuerzo.

Es obvio que la postura intencional funciona efectivamente cuando la meta es predecir el comportamiento de una computadora que juega al ajedrez, pues el propósito *para el que se diseñó* es «razonar» acerca de cuál es la mejor jugada o movimiento posible en el marco altamente racionalista del ajedrez. Si un programa de computadora dirige una refinería de petróleo, es casi igual de obvio que sus diversos movimientos se harán en respuesta a su detección de las condiciones que más o menos indican qué debería hacerse, dados los más amplios propósitos para los que fue diseñado. Aquí la presunción de *excelencia o racionalidad del diseño* resalta vivamente, pues el empeño de un programador incompetente podría dar lugar a un programa que rara vez hiciera lo que los expertos dicen que debería hacer en esas circunstancias. Cuando los sistemas de información (o de control) están bien diseñados, los fundamentos de sus acciones se reconocen de inmediato y son altamente predecibles —sin importar si los ingenieros que escribieron los programas siguieron la recomendable práctica de incluir «comentarios» al código fuente para explicar estos fundamentos a los curiosos— (más adelante abundaremos en estos temas). No tendríamos que saber nada de programación para predecir el comportamiento del sistema; necesitamos saber algo sobre las exigencias racionales que implica dirigir una refinería de petróleo.

Y ahora podemos ver por qué y cómo la postura intencional funciona para predecirnos a *nosotros*. Cuando nos tratamos *el uno al otro* como sistemas intencionales, de manera similar suavizamos nuestra ignorancia de los detalles de los procesos que ocurren dentro del cráneo del otro (¡y dentro del nuestro propio!) e inconscientemente nos fiamos del hecho de que, en una notablemente buena primera aproximación, la gente es racional. Si de repente nos vemos metidos en un escenario humano novedoso, normalmente podemos encontrarle un sentido sin esfuerzo, involuntariamente de hecho, gracias a nuestra capacidad innata de ver lo que la gente debería creer (la verdad acerca de lo que se pone frente a ella) y lo que debería desear (lo que es bueno para ella).

La fecundidad de nuestras previsiones basadas en la psicología vulgar no es objeto de polémica, pero sí hay un gran desacuerdo sobre cómo explicar esta munificencia. ¿Aprendemos decenas, cientos o miles de «leyes de la naturaleza» al estilo de

> «Si una persona está despierta y tiene los ojos abiertos y está frente a un camión, tenderá a creer que enfrente de ella hay un camión», y
> «Si la gente cree que cooperar con los demás puede traerle beneficios a un bajo

costo, tenderá a hacerlo, aunque no los conozca»,

o se trata de leyes en bruto generadas bajo demanda por un sentido implícito que nos dice que éstas son las respuestas racionales en esas circunstancias? Yo sostengo esto último. Si bien es verdad que hay muchos hábitos estereotípicos que pueden encapsularse en tales generalizaciones (que en principio podrían aprenderse una por una a lo largo de la vida), de hecho es difícil generar un escenario de ciencia ficción tan original, tan diferente de todos los predicamentos humanos, que la gente simplemente sea incapaz de imaginar cómo podría comportarse en esas circunstancias. «*¿Tú* qué harías si eso te pasara a ti?» es la pregunta que naturalmente se plantearía, y junto con respuestas tan poco útiles como «Probablemente me daría un síncope» viene la respuesta reveladoramente *racional:* «Bueno, espero ser lo bastante listo para ver que debería hacer X». Y cuando vemos a los personajes comportándose tan hábilmente en estos escenarios notablemente no estereotípicos, no nos cuesta trabajo entender qué están haciendo y por qué. Al igual que nuestra capacidad de entender oraciones totalmente novedosas de nuestros lenguajes naturales, oraciones que nunca antes en la vida hemos oído, nuestra capacidad de encontrar sentido en el inmenso despliegue de interacciones humanas denota una capacidad generativa que es hasta cierto punto innata en la gente normal.

Tan maquinalmente y con la misma naturalidad extendemos la postura intencional a los animales, lo cual es una táctica no optativa si estamos tratando de atrapar a una bestia astuta, y una táctica útil si estamos tratando de organizar nuestra comprensión de los comportamientos de animales más simples e incluso de plantas. La almeja tiene sus comportamientos, y son racionales dada su limitada perspectiva del mundo. No nos sorprende enterarnos de que los árboles que pueden percibir la lenta invasión de los rivales (porque ahora una mayor parte de la luz solar que cae sobre ellos se refleja en cosas verdes altas en los alrededores) reasignan los recursos para crecer más rápido. Después de todo, es lo más sensato para una planta en esas circunstancias. Entre los artefactos, hasta el humilde termostato aguanta una interpretación rudimentaria desde la postura intencional: quiere mantener la temperatura al nivel que le pediste, hace pruebas frecuentes de la temperatura para llegar a una creencia actualizada de la temperatura real, la compara con la temperatura deseada y actúa en consecuencia. Esto es lo que podrías decirle a un niño si le quieres explicar para qué sirve un termostato sin ponerte técnico.

Esta sencilla teoría de los sistemas intencionales trata de cómo y por qué encontramos sentido en los comportamientos de tantas cosas complicadas si las consideramos *agentes*. *No* es exactamente una teoría de los mecanismos internos que de algún modo logran la orientación racional así predicha. La postura intencional te da las especificaciones, la descripción del puesto, de un sistema intencional —por ejemplo, lo que debería discriminar, recordar y hacer— y deja que los ingenieros (o la evolución y el

desarrollo, cuando el sistema intencional sea un organismo) ejecuten esas especificaciones. Denme un agente que sepa reconocer entre un billete de un dólar y uno de diez dólares, calcular el cambio, detectar billetes falsos, y pueda y quiera entregar el producto que desea el consumidor 24 horas al día. Esta caracterización de la postura intencional es la especificación de una máquina expendedora o bien una rudimentaria descripción del puesto de dependiente de una tienda de abarrotes abierta día y noche, y no dice absolutamente nada sobre las entrañas de la entidad o las habilidades que pueda tener.

Esta equivalencia, o neutralidad, no es un «virus» sino una característica, como les gusta decir a los ingenieros de *software:* deja que la teoría de los sistemas intencionales sea la principal organizadora a la hora de salvar el abismo de confusión entre nuestras mentes y nuestros cerebros, como veremos en los siguientes tres apartados. Dicho brevemente, nos permite ver *qué tienen en común* las creencias «reales» (las creencias de *personas*) y los «simples» estados como de creencia (de las máquinas expendedoras, los animales, los niños pequeños y, algo utilísimo, partes *subpersonales* de las personas). Con el fin de prever, podemos usar la postura intencional para dar las especificaciones de las *competencias* de los subsistemas del cerebro antes que cualquier conocimiento detallado de cómo se ejecutan esas competencias. Analizamos a la persona «real», grande y elaborada, hasta convertirla en agentes subpersonales, con sus propias prioridades y métodos, y luego analizamos éstos a su vez hasta convertirlos en agentes todavía más simples y tontos. A la larga llegamos a sistemas intencionales que son lo suficientemente simples para describirlos sin más ayuda de la postura intencional. Salvar el abismo entre la **psicología vulgar al nivel de la persona** y las actividades **subpersonales** de los circuitos neuronales es una tarea pasmosa de imaginación que saca provecho de esta relajación por principio de las condiciones que los filósofos han tratado de imponer a la creencia y el deseo humanos genuinos y adultos. ¿En qué lugar de la cuesta descendente hacia la coseidad insensata se detienen los actos «reales» de creer y desear, y los relevan los simples «como si» se creyera y «como si» se deseara? Esta exigencia de una línea diáfana está mal motivada, como vimos en el apartado 15 y volveremos a ver con la ayuda de varias otras herramientas de pensamiento.

El uso de la postura intencional tanto en la informática como en la psicología animal es ubicuo, y la teoría de los sistemas intencionales explica por qué. Algunos teóricos de la biología evolutiva dicen prescindir de ella, pero se engañan, como veremos en la sección dedicada a la evolución.[18]

19. La distinción personal/subpersonal

Tus ojos no ven: eres tú quien lo hace. Tu boca no disfruta el pastel de chocolate, sino tú. Tu cerebro no detesta el ardor en tu hombro, sino tú. Tu mano no firma un contrato, sino tú. Tu cuerpo podrá excitarse, pero tú te enamoras. No se trata sólo de un aspecto «gramatical», como el hecho de que cuando hay una tormenta decimos «Está lloviendo» y no «La tormenta está lloviendo». Tampoco se trata *sólo* de la convención que son las definiciones de las palabras. La gente a veces pregunta en tono displicente: «¿Qué no es pura semántica?», lo que se entiende como que pocas cosas dependen de cómo «definamos los términos». No obstante, a menudo importa mucho cómo definamos los términos, y ésta es una de esas veces. Nuestra manera de hablar de las *personas* y de lo que pueden hacer y padecer se funda en algunos hechos importantes.

A primera vista parece que hay cosas que puede hacer una persona entera y que ninguna de sus partes propiamente dichas pueden hacer, y esto es casi cierto, pero un macabro ejercicio de imaginación sugiere que si es así, entonces una persona es (a grandes rasgos) una parte de un cuerpo humano propiamente dicha, a saber, un cerebro intacto que funciona. (¿*Tienes* un cerebro o *eres* un cerebro? No es una pregunta fácil de responder.) Si me cortas los brazos, aun así puedo firmar un contrato (con una pluma entre los dedos de los pies o una directriz vocal), pero si apagas mi cerebro, nada que mis brazos y manos puedan hacer cuenta como firmar un contrato. Sácame los ojos y no podré ver, a menos que me coloquen unas prótesis a modo de globos oculares, lo cual no es una remota fantasía de ciencia ficción. ¿Y si empiezas a «amputar» partes de mi cerebro? Si quitaras la corteza occipital pero dejaras intactos mis ojos y el nervio óptico, tendría «ceguera cortical», pero mantendría alguna visión residual (como en la famosa enfermedad conocida como visión ciega). Sin duda podríamos amputar un poco más de cerebro y acabar con esta visión residual, y *tú* todavía podrías llevar una vida. La idea tentadora es que un proceso así de eliminación, que quita el oído, el tacto, el gusto y el olfato, podría reducir el cerebro a los principales cuarteles generales de *ti:* y es allí donde una persona estaría, y es eso lo que una persona sería. Tentadora pero equivocada. Las innumerables competencias del cerebro están tan entrelazadas e interactúan tanto entre sí que simplemente no hay un sitio central en el cerebro «donde todo se reúna» para la conciencia.[19] En realidad, muchas de las competencias, disposiciones, preferencias y peculiaridades que te hacen ser *tú* dependen de caminos que pasan por tu cuerpo fuera de tu cerebro; el popularísimo experimento mental filosófico del trasplante del cerebro (¿quién preferirías ser: el «donante» del cerebro o su «receptor»?) es posible gracias a una idealización muy distorsionada. Como dije en una ocasión, «no se me puede arrancar de mi cuerpo dejando un corte limpiecito».[20]

Probablemente la propiedad más importante de internet sea su descentralización; en

ninguna parte del mundo tiene un centro o un cuartel general en el que soltar una bomba certera pudiera terminar con él. Sus partes propiamente dichas tienen un alto grado de redundancia y versatilidad, así que se «degrada elegantemente», si acaso se degrada, cuando cualquier parte suya se inhabilita. Hal, la computadora inteligente de *2001: Odisea del espacio,* tiene un «Centro de Memoria Lógica», una habitación llena de casetes de memoria que Dave puede desmontar uno por uno, hasta apagar a Hal por completo. Internet no tiene un centro así, y si bien la naturaleza no nos ha equipado con tan magnífico nivel de invulnerabilidad distribuida, aún hay una considerable descentralización de *ti* en tu cuerpo, así como una considerable versatilidad en tus partes. La organización de tu cerebro demuestra ser extraordinariamente *plástica* (capaz de transformarse en nuevas configuraciones), de modo que puedes seguir siendo tú, perseguir tus sueños, confundir a tus enemigos, tramar tus estratagemas, revivir tus tribulaciones y tus triunfos, a pesar de que te hayan extirpado partes del cuerpo importantes pero no del todo «vitales». Ésa es una de las razones por las que *tú* puedes tener competencias que ninguna de tus partes tiene. O podemos voltear al revés el asunto: la única manera de entender o interpretar los poderes de las *partes* de un cuerpo vivo entero es considerar cómo contribuyen a la coordinación del magnífico sistema en su totalidad.

Piensa en estos otros ejemplos: no es tu cerebro el que entiende español, sino tú. A tu cerebro no le parece gracioso el chiste, sino a ti. Incluso si las actividades de las estructuras competentes de tu cerebro desempeñan los principales papeles causales en tu entendimiento y en tu apreciación, no podrían hacer su trabajo sin el adiestramiento y el apoyo que a lo largo de los años proporcionan tus órganos de los sentidos, tus miembros y otros efectores.[21]

Así, no es sólo una convención postular a una *persona,* el agente duradero, consciente y racional cuyo cuerpo vivo es éste, como el *sujeto del historial* de la mayoría de nuestras atribuciones cotidianas: eres *tú* quien cometió el error, ganó la carrera, se enamoró de Leslie, habla un francés aceptable, quiere ir a Brasil, prefiere a las personas rubias, profirió la calumnia (véase también el apartado 62). Tienes hambre, estás cansado y malhumorado, todo gracias a tus partes subpersonales, y nada más.

¿Pero entonces qué hay de esas partes propiamente dichas? ¿Son sólo como ladrillos apilados en un cuerpo humano viviente? La respuesta es sí, si vemos las partes más pequeñas: los átomos, pero es no en todos los otros niveles, desde el molecular al celular y los más altos. Las proteínas que son las bestias de carga dentro de tus células son pequeños robots increíblemente competentes y sagaces: podríamos llamarlos nanorrobots. Las neuronas que hacen la mayor parte del trabajo de transmisión, control y regulación en tu cerebro son robots más versátiles y competentes: podríamos llamarlos microrrobots. Forman coaliciones que participan y cooperan en estructuras más grandes,

se comunican de un lado a otro, se reprimen entre sí, analizan el flujo de información que viene de los sentidos, despiertan estructuras de información latentes «en la memoria» (que no es un lugar del cerebro aparte) y organizan las sutiles cascadas de señales que mueven tus músculos cuando actúas.

Todos estos niveles más altos que los componentes atómicos básicos presentan un grado de *albedrío.* En otras palabras, se pueden interpretar como *sistemas intencionales.* En el plano molecular (las proteínas motoras, las enzimas correctoras en el ADN, los guardianes de billones de portales en las membranas de tus células, y otros por el estilo), sus competencias son muy «robóticas» sin dejar de ser imponentes, como los ejércitos de escobas que marchan en aquella escena de «El aprendiz de brujo» en la película *Fantasía,* o el demonio de Maxwell, por mencionar dos ejemplos ficticios. En el plano celular, las neuronas individuales tienen un comportamiento más exploratorio: husmean en busca de mejores conexiones, cambian sus modalidades de disparo como función de su experiencia reciente. Son como prisioneros o esclavos, más que simples máquinas (como las proteínas nanorrobots); podrías pensar en ellas como células nerviosas en celdas de la cárcel, trabajando miopemente en proyectos masivos de los que no tienen ni la menor idea, pero siempre ansiosas por mejorar su suerte cambiando de táctica. A niveles más altos, la miopía empieza a disiparse, mientras grupos de células —tractos, columnas, ganglios, «núcleos»— asumen funciones especializadas que son sensibles a las condiciones cada vez más diversas, incluidas las condiciones del mundo externo. Aquí el sentido de agencia es todavía más fuerte, pues los «trabajos hechos» requieren un *criterio* considerable y también *toma de decisiones.*

Estos agentes son como trabajadores de cuello blanco, analistas y ejecutivos con responsabilidades particulares, pero también, como los trabajadores de cuello blanco en todas partes, tienen una saludable dosis de celo competitivo y una disposición de apropiarse de cualquier poder con que se encuentren en el curso de sus actividades, e incluso usurpar el control de cualquier actividad mal defendida en la que estén ocupados sus vecinos u otros que estén en comunicación con ellos. Cuando llegamos a agentes en este nivel de competencia, las partes subpersonales son ciertamente ladrillos inteligentes, y podemos empezar a ver, al menos en líneas generales, cómo podríamos fabricar con ellos a toda una persona que comprende. («Es necesario algún ensamblaje», como se advierte en la caja que contiene todas las partes de la bicicleta, pero al menos no tenemos que cortar y doblar el metal ni encargarnos de todos los aspectos prácticos.)

Esta idea, que podemos dividir y conquistar el abrumador problema de imaginar cómo podría componerse una persona a partir de (nada más que) moléculas sin mentes, puede verse de abajo arriba, como acabamos de hacer, o de arriba abajo, empezando con la persona entera y preguntando qué colección más bien pequeña de homúnculos muy listos podría conjurarse para hacer todos los trabajos que se necesitan a fin de mantener

a una persona en funcionamiento. Platón fue pionero del enfoque de arriba abajo. Su análisis del alma en tres partes con apariencia de agentes, análogas a los Guardianes, los Auxiliares y los Trabajadores, o lo racional, lo irascible y lo concupiscible, no era un muy buen comienzo, por razones que se han estudiado bien a lo largo de los últimos dos milenios. El ello, el yo y el superyó de Freud del último siglo supusieron cierta mejora, pero la empresa de descomponer una mente entera en submentes empezó a tomar realmente forma con la invención de la computadora y el nacimiento del campo de la inteligencia artificial, que al principio tenía el objetivo explícito de analizar las competencias cognitivas de una persona completa (adulta, consciente, usuaria del lenguaje) y descomponerlas en una amplia red de especialistas subpersonales, como el generador de objetivos, el buscador de memoria, el evaluador de planes, el analista de percepciones, el analizador sintáctico, etcétera.

20. Una cascada de homúnculos

En la búsqueda milenaria por entender la mente, los teóricos a menudo han sucumbido a la tentación de imaginar un agente interno, un pequeño hombre —*homunculus* en latín— que se sienta en el centro de operaciones del cerebro y hace todo el trabajo inteligente. Si imaginas el sistema nervioso humano como, digamos, una inmensa red de conmutadores telefónicos (tal como les gustaba a los pensadores en épocas tan recientes como las décadas de 1950 y 1960), tienes el problema del operador en el centro: ¿está su mente compuesta de una red telefónica más pequeña con su propio operador central, cuya mente a su vez está compuesta de…? Cualquier teoría que proponga como principio un homúnculo central como ése está condenada a la perspectiva de una regresión infinita.

Sin embargo, a lo mejor el error no es postular un homúnculo, sino postular un homúnculo *central.* En mi primer libro, *Content and Consciousness,* cometí un error garrafal, tentado por una broma a la que no me pude resistir. Escribí:

> El «hombrecito en el cerebro», el «fantasma en la máquina» de Ryle, es una célebre no solución a los problemas de la mente, y aunque no es totalmente imposible que la analogía de la «escritura del cerebro» tenga alguna aplicación útil, sí parece que se limita a remplazar el hombrecito en el cerebro con un comité.[22]

¿Y exactamente qué hay de malo con un comité? (¡*Ajá*!: llamada a mi implícita reducción al absurdo.) Al cabo de un tiempo me di cuenta (en *Brainstorms*) de que la idea de remplazar al hombrecito en el cerebro con un comité no es tan mala; pienso que es una de las buenas ideas fundamentales de la ciencia cognitiva. Ésta era la clásica estrategia de la *inteligencia artificial a la antigua* (llamada GOFAI por las iniciales de la expresión en inglés) que llegó a ser conocida como *funcionalismo homuncular:*

> El programador de inteligencia artificial comienza con un problema caracterizado intencionalmente, y en consecuencia ve a la computadora de manera francamente antropomórfica: si *resuelve* el problema dirá que ha diseñado una computadora que puede *[e. g.]* entender preguntas en inglés. Su primero y más alto nivel de diseño descompone a la computadora en subsistemas, a cada uno de los cuales se le asignan tareas caracterizadas intencionalmente; compone un organigrama de evaluadores, recordadores, discernidores, supervisores, y así por el estilo. Se trata ciertamente de *homúnculos* […] Cada homúnculo a su vez se analiza en homúnculos *más pequeños,* pero, y esto es más importante, también *menos listos.* Cuando se alcanza el nivel en que los homúnculos no son más que sumadores y restadores, para cuando sólo necesiten la inteligencia para escoger el mayor entre dos números si se les ordena

hacerlo, habrán sido reducidos a funcionarios *que pueden ser remplazados por una máquina.*[23]

La ventaja particular de esta estrategia es que le movió el tapete a la objeción de la *regresión infinita.* De acuerdo con el funcionalismo homuncular, la ominosa regresión infinita puede esquivarse y remplazarse con una regresión *finita* que termina, como acaba de observarse, en operadores cuya tarea es tan sosa que pueden ser remplazados por máquinas. La clave para comprender mejor esto fue descomponer todo el trabajo que imaginábamos que el operador central hacía y distribuirlo entre agentes menores y más tontos cuyo trabajo a su vez era distribuido, y así sucesivamente.

Esto fue un buen avance, ¡pero un producto no deseado del enfoque de arriba abajo de la clásica inteligencia artificial a la antigua era su rígida eficiencia burocrática! Es posible que hayamos podido pensar en librarnos del rey central o del presidente de la compañía, pero seguíamos teniendo un ejército de mandos intermedios bajo las órdenes de un batallón de vicepresidentes (cuyas interacciones constituían el más alto nivel del sistema) y que mandan a sus subordinados a entrar en combate, que a su vez llaman a filas a más oficinistas de menor categoría, y así sucesivamente. Esta organización hipereficiente, sin movimientos improductivos, sin exceso de personal, sin insubordinación, en gran parte fue impuesto por el hecho de que las grandes computadoras que sirvieron de ejemplo a los primeros modelos de inteligencia artificial eran lentas y poca cosa para los estándares actuales, y la gente quería resultados rápidamente. Si quieres causarle una buena impresión al organismo de financiamiento, más vale que tu inteligencia artificial no se tarde horas en responder una simple pregunta. Debería ser muy eficiente en su ejecución. Además, escribir miles de líneas de código es mucho trabajo, y si has conseguido descomponer la tarea que tienes como objetivo —responder preguntas acerca de las rocas de la Luna, diagnosticar enfermedades del riñón o jugar al ajedrez, por ejemplo— en un conjunto manejable de subtareas que puedes tratar de programar y luego integrar en un sistema viable, obtendrás tu «prueba de concepto»[24] a un precio asequible en cuanto a tiempo y dinero.

Observa que las computadoras siempre se han diseñado para mantener las necesidades y el desempeño casi completamente independientes entre sí. En el *hardware,* la energía eléctrica se reparte equitativa y abundantemente; no hay riesgo de privar a ningún circuito. En cuanto al *software,* un benevolente sistema planeador reparte ciclos máquina a cualquier proceso que tenga la prioridad, y aunque puede haber una u otra clase de mecanismo de licitación que determine qué procesos son los prioritarios, ésta es una fila ordenada, no una lucha por la supervivencia. Como diría Marx, «de cada quien según sus capacidades, a cada quien según sus necesidades». El informático Eric Baum ha llamado acertadamente a esta jerarquía un control de «Politburó». Tal vez una vaga

comprensión de este tipo de control haya dado origen a la común intuición popular de que a una computadora nunca podría *preocuparle* nada. No porque esté hecha de los materiales equivocados —¿por qué sería el silicio un sustrato menos apropiado para las preocupaciones que el carbono?— sino porque su economía interna no tiene incorporados riesgos ni oportunidades, así que no *tiene* que preocuparse.

Las neuronas no son así. La generalidad de las células que componen nuestros cuerpos probablemente no son sino esclavos obedientes: un poco como las desinteresadas y estériles hormigas obreras en una colonia, haciendo trabajos estereotípicos y viviendo en un ambiente relativamente no competitivo (marxista). Las neuronas, sin embargo —ahora lo pienso—, deben competir vigorosamente en un mercado. ¿Para qué? ¿Qué podría *querer* una neurona? La energía y las materias primas para seguir prosperando, tal como sus ancestros los eucariontes unicelulares y los primos más lejanos, las bacterias y las arqueas. Las neuronas son una especie de robots biológicos; desde luego que no son conscientes en ningún sentido sustancioso. Recuerda que son células eucariotas, cercanas a las células de la levadura o de los hongos. Si las neuronas individuales son conscientes, ¡entonces también lo es el pie de atleta! Sin embargo, las neuronas, como sus primos unicelulares sin mente, son agentes altamente competentes en una lucha de vida o muerte, no en el medio ambiente que hay entre los dedos de tus pies, sino en el demandante medio ambiente del cerebro, donde se llevan la victoria las células que pueden conectarse más efectivamente, contribuir a las tendencias más influyentes en los niveles de las **máquinas virtuales** donde se distinguen los propósitos e impulsos humanos a gran escala. Muchos subsistemas del sistema nervioso están organizados como *procesos oponentes,* ocupados en un juego de estira y afloja entre dos subsistemas, cada uno de los cuales trata de salirse con la suya (nuestras emociones, por ejemplo, bien pueden verse como tormentas rivales desplazándose las unas a las otras lo mejor que pueden, interponiéndose frente a las otras o colaborando contra una tormenta más). Entonces ahora pienso que las dinámicas de procesos oponentes de las emociones, y las funciones que desempeñan al dominar nuestras mentes, se sustentan en una *economía* de la neuroquímica que aprovecha los talentos competitivos de las neuronas individuales. (Obsérvese cómo lo que se afirma es que las neuronas siguen siendo buenas en el trabajo en equipo dentro de la economía más global, a diferencia de los agentes más radicalmente egoístas, las células cancerosas. Recuerdo la máxima del biólogo ganador del Nobel, François Jacob, de que el sueño de toda célula es convertirse en dos células; las neuronas disputan entre sí por mantenerse activas y ser influyentes, pero no sueñan con multiplicarse.) Desde este punto de vista, el dominio inteligente del comportamiento de un animal sigue siendo un proceso informático —tal como una transacción en el mercado bursátil es un proceso informático—, pero las neuronas son *neuronas egoístas,* como el neurocientífico Sebastian Seung ha dicho, que

procuran maximizar sus entradas de los distintos valores de recompensa que hemos encontrado en el cerebro.[25] ¿Y que *compran* las neuronas con su dopamina, serotonina u oxitocina? Adquieren mayor influencia en las redes en que participan, y en consecuencia una mayor seguridad (el hecho de que las mulas sean estériles no les impide velar por sí mismas, del mismo modo que las neuronas pueden moverse por instintos de autoprotección que heredaron en última instancia de sus ancestros reproductores).

Así, el funcionalismo homuncular de abajo arriba, inspirado en las neurociencias, se ve cada vez más como un modelo de cómo funciona un cerebro, ya que las «arquitecturas informáticas» más caóticas y competitivas que genera son más viables desde un punto de vista biológico: podemos empezar a distinguir los procesos de desarrollo que pudieron construir y revisar estas arquitecturas, empezando en el embrión y siguiendo hasta la adultez, y también ver cómo podrían haber evolucionado a partir de sistemas nerviosos más simples, que a su vez son equipos de homúnculos menos consumados, que maomeno perciben, señalan y recuerdan.[26]

21. El operador *maomeno*

¿Por qué permitirnos *maomeno* hablar de esto? Porque cuando analizamos —o sintetizamos— esta pila de niveles cada vez más competentes necesitamos seguir la pista de dos hechos acerca de cada nivel: qué *es* y qué *hace.* Qué *es* puede describirse desde el punto de vista de la organización estructural de las partes de que está hecho, mientras podamos asumir que las partes funcionan como deberían. Qué *hace* es alguna función (cognitiva) que *(maomeno)* realiza: lo suficientemente bien para que en el nivel que le sigue hacia arriba podamos suponer que en nuestras existencias hay una pieza competente que realiza exactamente esa función —*maomeno,* lo bastante buena para usarse—. Ésta es la clave para resolver la parte más difícil de la inconcebiblemente compleja cuestión de cómo podría jamás una mente estar compuesta de mecanismos materiales. En los albores de la era de la computación, Alan Turing, que si alguien merece el crédito de inventor de la computadora es él, vio esta posibilidad. Él podía empezar con pedazos de mecanismos carentes de la menor traza de mentalidad y organizarlos en mecanismos más competentes, que a su vez podían organizarse en mecanismos aún más competentes, y así sucesivamente, sin un límite claro. Lo que podríamos llamar el operador *maomeno* es, en ciencia cognitiva, análogo al gradualismo en procesos evolutivos de la teoría de Darwin (en el capítulo vi se habla más de esto). Antes de que hubiera bacterias había *maomeno* bacterias, antes de que hubiera mamíferos había *maomeno* mamíferos, y antes de que hubiera perros había *maomeno* perros, etcétera.

Necesitamos el gradualismo de Darwin para explicar la enorme diferencia entre un mandril y una manzana, y necesitamos el gradualismo de Turing para explicar la enorme diferencia entre un robot humanoide y una calculadora de bolsillo. El mandril y la manzana están hechos de los mismos ingredientes básicos, estructurados de manera diferente y explotados en una cascada de muchos niveles de diferentes competencias funcionales. No hay una línea divisoria por principio entre un *maomeno* mandril y un mandril. Tanto el robot humanoide como la calculadora de bolsillo están hechos de los mismos ladrillos básicos de Turing, que ni piensan ni sienten, pero conforme armamos con ellos estructuras más grandes y más competentes, que entonces se vuelven los elementos de estructuras aún más competentes en niveles más altos, a la larga llegamos a partes tan *(maomeno)* inteligentes que pueden ensamblarse en competencias que merecen llamarse comprensivas. Usamos la postura intencional para seguir la pista de las creencias y los deseos (o «creencias» y «deseos», o *maomeno* creencias y *maomeno* deseos) de los agentes *(maomeno)* racionales en todos los niveles, desde la simple bacteria, a través de todos los circuitos discernidores, señaladores, comparadores, recordadores que abarcan los cerebros de los animales desde la estrella de mar hasta los astrónomos. No hay reglas para trazar una línea arriba de la cual se encuentre la verdadera comprensión,

incluso en nuestro caso. La niña pequeña *maomeno* entiende su propia oración «Mi papá es doctor», y yo *maomeno* entiendo «$E=mc^2$».

Algunos filósofos se oponen a este *antiesencialismo* (véase el apartado 43): o bien crees que la nieve es blanca o no lo crees; o bien eres consciente o no lo eres; nada cuenta como aproximación de ningún fenómeno mental; es todo o nada. Y para esos pensadores, los poderes de las mentes son misterios insolubles porque las mentes son «perfectas», y perfectamente distintas de cualquier cosa que pueda encontrarse en meros mecanismos materiales.

22. Tejido maravilloso

En su estupendo libro sobre la magia callejera en la India, *Net of Magic: Wonders and Deceptions in India,* Lee Siegel escribe:

> «Estoy escribiendo un libro sobre magia», explico, y me preguntan: «¿Magia de verdad?» Por *magia de verdad* la gente se refiere a los milagros, los actos taumatúrgicos y poderes sobrenaturales. «No —respondo—: trucos de magia.» En otras palabras, la *magia de verdad* significa la magia que no es de verdad, mientras que la magia que es de verdad y que de hecho puede hacerse *no es magia de verdad.*[27]

La «magia de verdad» es —podría decirse que por definición— milagrosa, una transgresión de las leyes de la naturaleza. Mucha gente quiere creer en la magia de verdad. Cuando James Randi, mago, escéptico y cazafantasmas supremo, duplica los trucos de los autodenominados psíquicos como Uri Geller, lo que hace es mostrar que los efectos asombrosos no son magia de verdad: son prestidigitación. A algunas personas, empero, esto no las convence. En la sesión de preguntas y respuestas que siguió a una función en Winnipeg hace muchos años, un miembro del público acusó a Randi de doble engaño: según esto él era un psíquico tan real como Geller, pero estaba aparentando ser un mero prestidigitador que desenmascaraba la duplicidad de Geller para poder montarse en su fama y fortuna. Es difícil rebatir esa extraña impugnación, salvo enseñando a todo el público exactamente cómo se hacen los trucos, algo a lo que Randi, fiel a la tradición de los magos en todo el mundo, es renuente (Penn y Teller han promovido rutinas que sí revelan los mecanismos secretos de los trucos de magia, y después de aguantar la repulsa de algunos colegas, han mostrado cómo el tabú tradicional puede transgredirse sin echar a perder el espectáculo de magia).

Parecidas ansias de creer en la magia de verdad aquejan a mucha gente cuando se trata de la relación entre la mente y el cerebro. Algunas personas, entre ellas no pocos neurocientíficos y psicólogos —y filósofos—, se ven al menos subliminalmente atraídos por la idea de que las propiedades dinámicas del tejido nervioso de una u otra manera pueden hacer algo que podría llamarse milagroso, algo que utiliza fuerzas ocultas nunca soñadas por la ciencia. A lo mejor tienen razón, pero no debemos suponer desde el principio que la tienen. La regla debe ser: ¡*nada de tejidos maravillosos!*

He aquí algo que sabemos con una certeza casi total: nada que sea físicamente inexplicable tiene un papel en ningún programa de computadora: ningún campo de fuerza inconcebible hasta este momento, ningún chanchullo cuántico misterioso, ningún *élan vital.* Con toda seguridad, ninguna computadora tiene tejido maravilloso. Sabemos

exactamente cómo se llevan a cabo las tareas básicas en las computadoras y cómo pueden formar parte de tareas más y más complejas, y podemos explicar estas competencias construidas sin un solo indicio de misterio. Así, aunque el virtuosismo de las computadoras actuales sigue asombrándonos, las computadoras mismas, en cuanto máquinas, son tan mundanas como los abrelatas. Mucha prestidigitación, pero nada de «magia de verdad».

Esto es un hecho valioso, tanto que en la siguiente sección tendrá una demostración detallada. Su valor reside en el hecho de que cada vez que podemos hacer que una computadora haga algo que ha *parecido* milagroso, tenemos la prueba de que *puede* hacerse sin tejido maravilloso. *Quizá* el cerebro lo haga de otra manera, quizá incluso con tejido maravilloso (¡*quizá* Randi sea un verdadero psíquico, igual que Geller!), pero no tenemos ninguna buena razón para creerlo. Las computadoras, por lo tanto, tienen un importante papel como desmitificadoras, y ésa es una buena razón para insistir en que se desarrollen modelos informáticos de cualquier cosa que estemos tratando de entender, sean huracanes, burbujas inmobiliarias, VIH o la conciencia humana.

La expresión *tejido maravilloso* es una herramienta de pensamiento parecida a la cachiporra de un policía: la usas para reprender, para disuadir a otros de ocuparse de teorizaciones ilícitas. E igual que una cachiporra, puede abusarse de ella. Es un accesorio especial de la herramienta de pensamiento conocida como la navaja de Ockham y en consecuencia impone cierto conservadurismo científico, que puede ser miope. Debo mi ejemplo favorito de esto a William Bateson, uno de los padres de la genética moderna; he aquí lo que tenía que decir hace no *tanto* tiempo, en 1916:

> Las propiedades de las cosas vivientes están de alguna manera sujetas a una base material, tal vez en algún grado especial a la cromatina nuclear [cromosomas]; y sin embargo es inconcebible que partículas de cromatina o de cualquier otra sustancia, por compleja que sea, puedan poseer esos poderes que deben atribuirse a nuestros factores o genes. La suposición de que partículas de cromatina, indistinguibles unas de otras y de hecho casi homogéneas según cualquier prueba conocida, puedan por su naturaleza material conferir todas las propiedades de la vida sobrepasa el ámbito de incluso el más convencido de los materialismos.[28]

Sencillamente no podía imaginar el ADN. La idea de que pudiera haber tres mil millones de pares de bases en una doble hélice dentro de toda célula humana era algo que simplemente no estaba al alcance de su imaginación. Afortunadamente, otros biólogos no compartían el pesimismo de Bateson, y llegaron a descubrir exactamente cómo la hazaña aparentemente milagrosa de transmitir información genética de generación en generación se logra gracias a unas moléculas de lo más fantásticas. Sin embargo, a lo largo de esa

senda de descubrimientos, se aferraron a la regla «nada de tejido maravilloso». Habían aprendido de la genética mucho sobre las competencias que debía tener su presa, y se impusieron la tarea de construir un modelo físicamente posible de algo que las tuviera.

Hoy en día enfrentamos una tarea similar. La psicología experimental nos está dando un catálogo cada vez más detallado de las competencias y flaquezas de la mente —los triunfos de la percepción y los bochornos de la ilusión, el ritmo del aprendizaje de la lengua y los estados de distracción, deseo, miedo y alegría— y ahora, como «materialistas convencidos», tenemos que entender cómo diablos hace todo eso el cerebro, sin postular tejido maravilloso.

Conforme aumenta nuestra comprensión, cambia lo que cuenta como tejido maravilloso. Cuando el «conexionismo» y otros modelos de «redes neuronales» irrumpieron en escena a mediados de la década de 1980, demostraron una capacidad de aprendizaje y unos poderes para reconocer regularidades que nadie se habría atrevido a postular para pequeños tractos de neuronas unos cuantos años antes.[29] Todavía no sabemos exactamente cómo el cerebro explota los poderes informáticos que presentan estos modelos semirrealistas (tampoco sabemos si los explota), pero ahora está bien postular una competencia conexionista para alguna red neuronal que *todavía* no puede explicarse, mientras se sea franco al respecto y la competencia no esté claramente más allá del ámbito demostrado de hazañas (puede ser que Randi no haga el truco exactamente de la misma manera que Geller, pero no es riesgoso concluir que hay en el método de Randi alguna variación que explica la aptitud de Geller y que nos orienta en la investigación de los procesos que realmente entran en juego). La principal objeción al tejido maravilloso es que no nos da una manera de solucionar el problema, sino una manera de renunciar al intento y de suponer que es un misterio que jamás podrá resolverse.

23. Atrapado en el centro de operaciones del robot

Los robots no tienen **tejido maravilloso** (por definición, de hecho), así que ofrecen una plataforma aséptica para experimentos mentales como éste:

Una mañana te despiertas para descubrir que estás en una cama ajena en un extraño cuarto sin ventanas; dos de sus paredes están cubiertas de pequeñas luces de distintos colores que se encienden y se apagan, y las otras dos están cubiertas de miles de botones. Las luces y los botones tienen números, pero no letreros. Alguien te dejó una nota en el buró:

> ¡Buenos días! Te drogaron y te secuestraron mientras dormías y te trajeron aquí, a tu nuevo hogar. Hay comida en el refrigerador y un baño en el rincón, así que todas tus necesidades corporales inmediatas serán satisfechas. Estás encerrado en el centro de operaciones de un robot gigante. Cada luz, cuando se enciende, ofrece información sustancial y pertinente sobre las circunstancias del robot; todas estas luces son las salidas de analizadores de redes neurales altamente sofisticados correspondientes a las entradas crudas que surgen de los ojos de video de alta definición, los oídos de micrófono y los sensores de tacto y olfato del robot. Los botones inician acciones robóticas, todas coordinadas y listas para ejecutarse.
>
> El robot habita en un mundo peligroso, con muchos riesgos y oportunidades. Su futuro está en tus manos y, desde luego, también el tuyo depende de cuánto éxito consigas al pilotar tu robot por el mundo. Si es destruido, se apagará la electricidad de este cuarto, dejará de haber comida en el refrigerador y morirás. ¡Buena suerte!

Es un predicamento horrible. Con el corazón en la boca empiezas a experimentar y aprietas botones para ver qué pasa. Aprietas el botón amarillo con el número 4328 y observas que cuando lo haces titila y se apaga. ¿Estabas rascando la comezón del robot o cerrando sus ojos o tal vez «comiendo» algo, y por tanto satisfaciendo sus apremiantes necesidades metabólicas? Aprietas por segunda vez el botón 4328 y ahora se encienden otras luces dispersas. ¿Qué ha cambiado en el mundo? ¿Qué *significa* esto? Te atormenta porque te dijeron que en esas luces hay una tremenda cantidad de *información,* pero ¿qué luz señala qué información, y qué botón ordena qué acción robótica?

¡Si tan sólo las luces y los botones tuvieran letreros! Entonces —si los letreros estuvieran en un idioma que entendieras— podrías resolver el problema. ¡O si tan sólo hubiera en el cuarto una ventana que pudieras abrir, para asomarte hacia fuera y ver qué se mueve cuando aprietas un botón! ¿Tiene el robot brazos y piernas? Con una ventana podrías tratar de correlacionar los sucesos del mundo externo con las luces intermitentes.

Sin una ventana tienes toda esta información al alcance de la mano, pero ningún modo de interpretarla. Tienes cientos de acciones robóticas que poner en marcha, pero ningún modo de entender qué efectos pueden conseguir en el mundo.

Parece que estás en una situación imposible. Por listo e imaginativo que seas, no podrás entender qué significan los sucesos que ocurren en las paredes de tu cuarto con todos los datos a tu disposición. No obstante, si no puedes resolver este problema tenemos algo que se acerca a una paradoja, ¡porque tu predicamento aquí no es más que el predicamento de tu cerebro! Está atrapado en un cuarto sin ventanas —tu cráneo— con millones de líneas de entrada verdaderamente bullendo de información sobre el mundo externo y tus circunstancias corporales, y millones de líneas de salida colocadas para esimular a los músculos y contraerlos o relajarlos. Y tu cerebro no puede abrir una ventana en tu cráneo y asomarse hacia fuera y ver qué está pasando para causar esas repeticiones de señales que entran a raudales a tu corteza visual (y en todo caso, ¿de qué le serviría a tu cerebro esa ventana? A diferencia de ti, no tiene ojos aparte de los ojos cuyas señales recibe y trata de interpretar, ni recuerdos previos de las apariencias de las cosas en el mundo).

Puedes pensar que a lo mejor la tarea del cerebro se facilita por el hecho de que el dibujo excitatorio —las luces intermitentes— en la superficie de la corteza visual cuando ves un pato, por ejemplo, ¡de hecho tiene forma de pato![30] Esto estaría bien si pudiéramos suponer que tu cerebro aprendió, como tú, qué apariencia tiene un pato, ¿pero cómo podría aprender eso tu cerebro?

¿Cómo puede tu cerebro aprender lo que sea si primero tiene que «descodificar» todas sus señales? ¿Y descodificarlas a qué? ¿Al turco? Los letreros en turco no significan nada para ti, a menos que entiendas turco. ¿Tu cerebro tiene que entender una lengua antes de poder encontrarles a sus entradas algún valor? Como ya hemos visto, esta idea, que los cerebros vienen equipados con un lenguaje interno que nunca tienen que aprender —el mentalés, o Lenguaje del Pensamiento[31]— es tentadora. Parece ser un paso adelante en la dirección correcta, pero hasta que no nos den los detalles sobre cómo funciona y cómo evolucionó en un comienzo declarar que existe tal lenguaje del pensamiento no es más que darle otro nombre al problema sin solucionarlo. Sabemos que el cerebro *de alguna manera* resuelve el problema; nuestros cerebros sí se las arreglan, de manera muy solvente, para encontrar las salidas apropiadas para hacer frente a los predicamentos y oportunidades que la información de entrada presagia. Y sabemos que la solución del cerebro —sea cual sea— no puede ser *exactamente* como un lenguaje (como el español o el turco), pues no se adquiere en la infancia de la manera como adquirimos nuestras lenguas maternas. ¿Será más como un lenguaje escrito (con recuerdos debidamente inscritos en los archivos cerebrales) o un lenguaje solamente hablado? ¿Tiene un vocabulario de miles, de millones o de billones de «palabras»?

¿Importa el orden de las palabras? ¿Hay una gramática? ¿Puede una parte del cerebro malinterpretar el mensaje de otra parte?

Si la hipótesis del lenguaje del pensamiento postula en el centro de operaciones a un homúnculo que entiende el lenguaje (como si tú en el robot gigante pudieras leer letreros en las luces y los botones), simplemente pospone la tarea de entender cómo el aprendizaje y la comprensión pueden estar compuestos de maquinaria que consiste en partes que no comprenden. Y si *no* postula un homúnculo que descodifica los mensajes, entonces el sistema, sea éste lo que sea, a fin de cuentas no se parece mucho a un lenguaje. Y como nadie ha encontrado los pormenores de cómo funcionaría tal lenguaje del pensamiento, o cómo se instalaría a lo largo del desarrollo y la experiencia, probablemente sea mejor no tomarnos el pelo, no hacernos creer que estamos progresando cuando probablemente no es así.

Esta bomba de intuición nos muestra que en un sentido importante tu cerebro *no* se encuentra en el mismo predicamento que tú, atrapado en el centro de operaciones. Su tarea —como debe ser— queda parcialmente resuelta desde antes por la manera como algunas entradas se «conectan» con algunas salidas de modo que en el cerebro haya algún apalancamiento que ayude a aprender y refinar otras relaciones apropiadas. Ésta es otra manera de exagerar la ampliamente reconocida afirmación de que cuando nacemos nuestro cerebro no es una tábula rasa,[32] sino que ya está diseñado por la selección natural para expresar diversas preferencias, previsiones y asociaciones. Y mientras estén incorporadas algunas de las conexiones apropiadas, no hace falta que tengan letreros.

Antes de que pueda haber comprensión, tiene que haber **competencia sin comprensión**. Así es la naturaleza. Las bacterias tienen toda clase de competencias extraordinarias que ellas no necesitan entender para nada; sus competencias les sirven mucho, pero ellas mismas no están enteradas. Los árboles tienen competencias que les proporcionan muchos beneficios si las ejercitan, pero no tienen que saber por qué. El proceso mismo de selección natural es divinamente competente, un generador de diseños de excepcional ingenio y eficacia, sin una pizca de comprensión.

La clase de comprensión de la que gozamos los adultos humanos es un fenómeno muy reciente en el escenario evolutivo, y tiene que estar compuesta de estructuras cuya competencia viene acompañada de una mínima clase de semicomprensión o seudocomprensión, que además la posibilita. Es la clase de (hemi-, semi-, demi-) comprensión de que gozan los peces o los gusanos. Estas estructuras están diseñadas para comportarse apropiadamente casi todo el tiempo, sin tener que saber por qué su comportamiento es apropiado.

La alternativa, poner a un comprendedor hecho y derecho en el centro de operaciones, enfrentado a todas las entradas y salidas, es un callejón sin salida garantizado. ¿Por qué? Porque si este poder de comprensión es inexplicable, has

instalado tejido maravilloso, un milagro, en la base de tu teoría, y si es explicable —desde el punto de vista de procesos, actividades y poderes que a su vez carecen de ese poder de comprensión—, desperdiciaste tu turno y estás otra vez en el principio, con todos los que estamos tratando de explicar cómo es que la comprensión nace de la competencia.

[1] Algunos entusiastas dirán que los animales sí hablan, en lenguajes que aún no descubrimos o traducimos. Es una idea perennemente atractiva, pero las investigaciones exhaustivas sobre la comunicación de simios, aves y delfines que se ha llevado a cabo recientemente muestran que la habilidad de los animales para «comunicar sus pensamientos» es muy limitada. Si eso es lenguaje, entonces las termitas son ingenieras, las mantis religiosas tienen una fe y los lobos tienen legislaturas.

[2] Jerry Fodor blandió muy efectivamente esta pregunta retórica en su libro pionero *The Language of Thought,* Harvester Press, Hassocks, 1975, p. 27, y subrayó la respuesta que quería insinuar citando el desafiante comentario de Lyndon Johnson a los estadunidenses: «Soy el único presidente que tienen».

[3] *Folk psychology;* en ocasiones también se traduce como «psicología popular» o «psicología del sentido común». En «psicología vulgar», que aquí se ha preferido, se emplea la acepción de «vulgo» como «el común de la gente popular» (Real Academia Española, *Diccionario de la lengua española*), bastante cercana a la de *folk* como «gente en general» *(Oxford Dictionary of English).* [T.]

[4] A esta conclusión se le suele llamar el *holismo* de lo mental (o de lo intencional); Jerry Fodor ha negado incondicionalmente el holismo: él dice que no le cuesta ningún trabajo imaginar a una criatura con exactamente una creencia. Véase J. Fodor y E. Lepore, *Holism: A Shopper's Guide,* Blackwell, Oxford, 1992.

[5] W. Sellars, «Philosophy and the Scientific Image of Man», en *Science, Perception and Reality,* Routledge & Kegan Paul, Londres, 1992, p. 1.

[6] J. von Uexküll, «A Stroll through the Worlds of Animals and Men: A Picture Book of Invisible Worlds», en C. H. Schiller (comp. y trad.), *Instinctive Behavior: The Development of a Modern Concept,* International Universities Press, Nueva York, 1957.

[7] J. J. Gibson, *The Ecological Approach to Visual Perception,* Houghton Mifflin, Boston, 1979.

[8] Nos detendremos en los detalles importantes de este ejemplo en el apartado 27.

[9] Me disculpo con los lectores a los que les guste fantasear sobre cómo sería ser una margarita, y les ofrezco de consolación el concepto de la *Umwelt:* permite hacer justicia a la muy real destreza discriminatoria de una margarita sin pasarse de la raya.

[10] W. C. Wimsatt, «Randomness and Perceived Randomness in Evolutionary Biology», *Synthese,* 43: 287-290, 1980.

[11] Este aforismo se atribuye a diferentes personas, entre ellas el filósofo Ludwig Wittgenstein, el pintor Paul Klee y el crítico Viktor Shklovski.

[12] J. Barrett, «Exploring the Natural Foundations of Religion», *Trends in Cognitive Science,* 4: 29-34, 2000; véase también D. C. Dennett, «Intentional Systems in Cognitive Ethology: The "Panglossian Paradigm" Defended», *Behavioral and Brain Sciences,* 6 (3): 343-390, 1983.

[13] D. C. Dennett, *Breaking the Spell: Religion as a Natural Phenomenon,* Viking Penguin, Nueva York, 2006.

[14] Dennet dedicó a este concepto su obra *The Intentional Stance,* traducida al español con el título *La actitud intencional.* Aquí se ha preferido traducir «stance» como «postura», pues esta palabra refleja mejor la idea de una posición que se adopta frente a un asunto moral o intelectual, en contraste con una posición del ánimo o del cuerpo, que son las principales asociaciones con «actitud». [t.]

[15] Las excepciones ayudan a apreciar el contraste. Uno de los síntomas más claros del autismo, tanto el grave como el leve, en sus diversas formas es la dificultad con la psicología vulgar. Las personas autistas que llegan a interpretar hasta cierto punto a otras personas, como Temple Grandin, en realidad *sí* parecen basarse en una teoría adquirida con dificultades. Donde a otros los vemos entender sin esfuerzo y, en efecto, involuntariamente los significados de los gestos, sonrisas y circunstancias, ella tiene que deducir estas implicaciones a partir de lo que observa. Hábilmente ha recogido indicios que le permiten identificar los sellos distintivos de un saludo amistoso, en contraposición a una amenaza, y reconocer cuando la gente está de acuerdo en algo. Ha aprendido por su cuenta a distinguir entre promesas y sobornos, chistes y mentiras. El personaje de Sheldon Cooper en *The*

Big Bang Theory es un conocido ejemplo ficticio de alguien con síndrome de Asperger, una forma leve de autismo, y vemos y oímos a su brillante mente científica en agitada acción, infiriendo los pormenores de las situaciones a las que se enfrenta como si estuviera analizando los efectos de los catalizadores de una reacción química en un tubo de ensayo. Está claro que Temple Grandin y el ficticio Sheldon Cooper tienen una teoría de la mente. Si los demás también tenemos una, no se parece mucho a la de ellos.

16 La llamé postura «intencional» en vez de, digamos, la postura del «agente racional» o la postura de la «creencia-deseo»porque en aquel momento (1971) gran parte del trabajo filosófico sobre la *intencionalidad* tenía que ver con la lógica de las «expresiones intencionales» que sirven para informar sobre las *actitudes proposicionales* de creencia, deseo, esperanza. Estas expresiones aluden a actitudes que son *acerca de* algo, y ese rasgo no sólo da lugar a problemas lógicos sino que también apunta a soluciones. Una creencia o una oración pueden ser acerca de algo que no existe, como Santa Claus, y un deseo puede ser de algo que no existe: la Fuente de la Eterna Juventud o simplemente un vaso de agua fría en medio del desierto. La postura intencional es *acerca* de aquello *acerca* de lo cual los agentes están pensando. *No* es algo que *intencionalmente* adoptemos en circunstancias normales; típicamente es involuntario y rara vez lo advertimos.

17 D. C. Dennett, *Elbow Room: The Varieties of Free Will Worth Wanting ,* The MIT Press, Cambridge, 1984.

18 Mi texto de 1983 «Intentional Systems in Cognitive Ethology: The "Panglossian Paradigm" Defended», que fue artículo central en *Behavioral and Brain Sciences,* ofrece los pormenores, acompañados de un coro fascinante de objeciones y refutaciones. Para una elaboración más reciente de la postura intencional y otra refutación a las objeciones comunes, véase mi ensayo «Intentional Systems Theory», en B. McLaughlin, A. Beckermann y S. Walter (comps.), *The Oxford Handbook of Philosophy of Mind,* Oxford University Press, Oxford, 2009, pp. 339-350, de donde provienen algunas partes de esta sección.

19 Ésta es la principal afirmación que defiendo en mi libro *Consciousness Explained* (Little Brown, Boston, 1991); la idea de un lugar especial así —lo llamo el Teatro Cartesiano— es por lo visto indómitamente tentadora, pues muchos pensadores todavía no la entienden, a pesar de todas las razones que otros autores y yo hemos presentado.

20 D. C. Dennett, *Kinds of Minds: Towards an Understanding of Consciousness,* Basic Books, Nueva York, 1996, p. 77.

21 No suele mencionarse que en ese experimento mental favorito de los filósofos, *el cerebro en la cubeta,* el cerebro viviente *debe* haber tenido una vida en un cuerpo para configurar todo el *hardware* con el *software* de años de experiencia. Si «sólo» tuvieras que construir un cerebro desde cero, dejando que las conexiones entre elementos fueran cualquier cosa que resultaran ser (en otras palabras, cablear al azar el flamante cerebro), la probabilidad de que tuviera alguna competencia o albergara ideas coherentes, intenciones, proyectos o recuerdos (aparentes), sería Evanescente (véase el apartado 35. Véase también el apartado 32).

22 D. C. Dennett, *Content and Consciousness,* Routledge & Kegan Paul, Londres, 1969, p. 87.

23 J. Haugeland, *Artificial Intelligence: The Very Idea,* The MIT Press, Cambridge, 1985, p. 80.

24 *Prueba de concepto* es una expresión de la jerga ingenieril útil pero peligrosa: si te piden que diseñes un artilugio muy sofisticado y complicado, diseñas un artilugio muy sencillo —resuelves lo que tiene el lindo nombre de *problema de juguete*— y, cuando funciona con su tarea simplificada, declaras que tienes una «prueba de concepto»; ahora «sólo» se trata de «agrandarla» al tamaño del artilugio del todo competente que te pidieron, que nada más requerirá más tiempo y dinero; el difícil problema conceptual ya se ha «resuelto». Esto a veces funciona tal como se anuncia.

25 S. Seung, «The Once and Future Science of Neural Networks», presentado en The Society for Neuroscience Meeting, San Diego, 4 de noviembre de 2007.

26 No a todo el mundo le gusta esta idea. Un libro reciente, *Philosophical Foundations of Neuroscience* (Wiley-Blackwell, Malden, 2003), resultado de la colaboración entre un neurocientífico, Max Bennett, y un filósofo, P. M. S. Hacker, la desdeña con ganas. Para una detallada refutación (y su refutación a mi refutación),

véase M. Bennett, D. C. Dennett, P. M. S. Hacker y J. Searle, *Neuroscience and Philosophy: Brain, Mind, and Language,* Columbia University Press, Nueva York, 2009.

27 L. Siegel, *Net of Magic: Wonders and Deceptions in India,* The University of Chicago Press, Chicago, 1991, p. 425.

28 W. Bateson, «Review of the *Mechanisms of Mendelian Heredity* by T. H. Morgan, A. H. Sturtevant, H. J. Muller and C. B. Bridges, New York, 1915», *Science,* 44: 536-543, 1916.

29 Para una visión de conjunto y una introducción, véase mi artículo «Logical Geography of Computational Approaches: A View from the East Pole», en R. M. Harnish y M. Brand (comps.), *The Representation of Knowledge and Belief,* University of Arizona Press, Tucson, 1986, pp. 59-79. Una inmejorable primera antología es J. McClelland, D. Rumelhart y The PDP Research Group, *Parallel Distributed Processing : Explorations in the Microstructure of Cognition,* vol. 2, The MIT Press, Cambridge, 1986.

30 Así es: los patrones excitatorios en la corteza occipital (visual) —y otras regiones— realmente adoptan las formas (distorsionadas) de los objetos del mundo que están frente a tus ojos, pero ¿hay algo en tu cerebro que esté *mirando* esas formas? Quizá algo maomeno ve esas formas. Hay que abordar esta idea con cautela.

31 J. Fodor, *The Language of Thought, op. cit.,* y *lot 2: The Language of Thought Revisited, op. cit.*

32 S. Pinker, *The Blank Slate: The Modern Denial of Human Nature,* Viking Press, Nueva York, 2002.

IV. Intermedio sobre computadoras

Tomemos un descanso. Como probablemente has podido apreciar, hasta aquí he mencionado varias veces las computadoras, y el tema nos dará todavía mucho de que hablar en las siguientes páginas. Las computadoras son, sin lugar a dudas, las herramientas para pensar más potentes que tenemos, no sólo porque les restan parte de talacha a gran cantidad de tareas intelectuales, sino porque muchos de los conceptos que han inventado los informáticos son herramientas de pensamiento excepcionales por derecho propio. En estos días todos estamos metidos hasta el cuello en una inundación de jerga informática —*software* y *hardware,* ancho de banda y gigahercios—, y seguramente muchos de los lectores han absorbido impresiones bastante exactas de lo que significan los nuevos tecnicismos de moda. He descubierto, sin embargo, que aunque mis estudiantes mueven la cabeza en señal de entendimiento cuando uso esos términos en mis clases, su comprensión es irregular, y a veces me confunden porque llegan a unas interpretaciones muy extrañas de lo que estoy tratando de comunicarles. Entonces aquí voy a enseñarle al lector cómo escribir programas para la computadora más simple del mundo.

Si ahora te tomas el tiempo y el esfuerzo de dominar algunas habilidades básicas, te redituará sustancialmente al profundizar en tu comprensión de lo que viene después (y si ya eres un experimentado profesional de la computación, algunas de mis maneras de explicar los temas podrán resultarte útiles cuando hables con los no expertos, o bien puedes saltarte esta sección). He hecho vuelos de prueba de los apartados siguientes con cientos de estudiantes de licenciatura que tienen fobia a las computadoras y en ellos logramos felices resultados: hasta los que preferirían memorizar una página del directorio telefónico a resolver un acertijo reconocen el placer, el cosquilleo de satisfacción, cuando consiguen que esta computadora estúpidamente simple haga lo que se espera de ella. Y cuando el lector complete el ejercicio, habrá sido iniciado en los siete secretos del poder de la computación.

24. Los siete secretos del poder de la computación revelados

Las computadoras tienen poderes que en siglos anteriores habrían parecido milagrosos —«magia de verdad»—, pero aunque muchos programas de computadora son abrumadoramente complicados, todos ellos están compuestos de pasos que son perfectamente explicables en términos muy sencillos. No hay lugar para el misterio en lo que hacen las computadoras. Ese mismísimo hecho es parte de su valor como herramientas de pensamiento, y explicar —a grandes rasgos— cómo funcionan es filosóficamente interesante de por sí. Bien vale la pena entender en un nivel elemental cómo hacen su «magia» las computadoras. Este apartado ofrece esa desmitificación.

Empezamos por imaginar la que probablemente sea la computadora más sencilla que pueda concebirse: la *máquina de registro,* sólo para ver cuáles son sus poderes y por qué. Luego veremos que una *máquina de Turing* y una *máquina de Von Neumann* (como tu computadora portátil) son como máquinas registradoras, sólo que más eficientes (cualquier cosa que pueda hacer tu computadora portátil puede hacerla una máquina de registro, pero mejor espera sentado pues puede tomarle siglos). Entonces entenderemos cómo podrían otras «arquitecturas» informáticas seguir multiplicando la velocidad y capacidad de nuestra máquina básica, la máquina de registro. La arquitectura del cerebro humano es, desde luego, una de las más interesantes e importantes arquitecturas que considerar.

Espera un momento. ¿Estoy diciendo que tu cerebro es sólo una computadora gigante? No; en todo caso, todavía no. Estoy señalando que *si* tu cerebro fuera una computadora gigante, *habría una manera* de entender todas sus actividades sin vestigios de misterio (ojalá pudiéramos encontrarla). Nuestro método será la *ingeniería inversa:* estudiar un sistema complicado para descubrir cómo hace lo que hace. La ingeniería inversa nos dice cómo el corazón ejecuta sus tareas de bomba y cómo los pulmones recogen oxígeno y expelen bióxido de carbono. La neurociencia es el empeño de hacer ingeniería inversa con el cerebro. Sabemos *para qué* sirven los cerebros: para prever, guiar, recordar y aprender; lo que nos hace falta es entender cómo consiguen todo esto.

Es un tema apasionadamente polémico. El novelista Tom Wolfe identificó el punto débil de las más encarnizadas polémicas con el título de su ensayo «Sorry, but Your Soul Just Died» («Lo siento, pero tu alma acaba de morir»).[1] Si vamos a explorar este territorio peligroso —y no nada más hacemos perder el tiempo a todos con declamaciones y denuncias— necesitaremos herramientas más afiladas. Necesitamos saber qué pueden hacer las computadoras *y cómo lo hacen* antes de poder abordar responsablemente la pregunta de si nuestros cerebros albergan y explotan fenómenos incomprensibles y milagrosos inalcanzables para todas las computadoras posibles. La

única manera satisfactoria de *demostrar* que tu cerebro no es —ni podría ser— una computadora sería mostrar *a)* que algunas de sus «partes móviles» realizan tipos de actividades de manejo de información que ninguna computadora podría realizar, o bien *b)* que las actividades simples que sus partes sí realizan no pueden componerse, agregarse, coordinarse, como hacen las computadoras, hasta constituir las hazañas mentales que conocemos y apreciamos.

Algunos expertos —no sólo filósofos sino neurocientíficos, psicólogos, lingüistas e incluso físicos— han argumentado que usar la «computadora como metáfora» del cerebro/mente humano es profundamente equívoco, y, de modo más drástico, que los cerebros pueden hacer cosas que las computadoras no. Generalmente, pero no siempre, estas críticas presuponen una concepción muy ingenua de lo que es o debe ser una computadora, y terminan demostrando la verdad obvia (e irrelevante) de que los cerebros pueden hacer muchas cosas que tu computadora portátil no (dada su magra dotación de transductores y efectores, su insignificante memoria, su límite de velocidad). Si queremos evaluar estas fuertes afirmaciones escépticas sobre los poderes de las computadoras *en general,* tenemos que entender de dónde viene el poder informático *en general* y cómo se lleva a cabo o cómo puede llevarse a cabo.

Quien en los albores de la era informática presentó la genial idea de una máquina de registro fue el lógico Hao Wang, alumno de Kurt Gödel, por cierto, y filósofo.[2] Es una elegante herramienta de pensamiento; debería formar parte de tu propia caja de herramientas. No es, ni de lejos, tan famosa como debería.[3] Una máquina de registro es una computadora idealizada, imaginaria (y perfectamente posible), que consta únicamente de (un número finito de) *registros* y una *unidad procesadora.*

Los *registros* son lugares de memoria, cada uno con una dirección única (registro 1, registro 2, registro 3, etc.), y cada uno puede tener, como *contenido,* un solo número entero (0, 1, 2, 3…). Piensa en cada registro como una gran caja que puede contener cualquier cantidad de frijoles, de 0 a…, según el tamaño que tenga. Normalmente consideramos que las cajas pueden tener como contenido *cualquier* entero, para lo cual desde luego se necesitarían cajas infinitamente grandes. Para nuestros propósitos, con cajas muy grandes nos basta.

La *unidad de procesamiento* está equipada con tan sólo tres sencillas competencias, tres «instrucciones» que puede «seguir» paso a paso, una por vez. Cualquier secuencia de estras instrucciones es un programa, y a cada instrucción se le da un nombre que la identifica. Las tres instrucciones son:

- *Finaliza.* Es decir, puede detenerse o apagarse.
- Registra un *incremento* a *n* (suma 1 al contenido del registro *n;* pon un frijol en la caja *n*) y ve a otro paso, el paso *m.*
- Registra una *disminución* a *n* (resta 1 del contenido del registro *n;* quita un

frijol de la caja *n*) y ve a otro paso, el paso *m*.

La instrucción de *disminuir* funciona igual que la instrucción de *incremento,* salvo por una sola complicación de enorme importancia: ¿qué debe hacer si el número del registro *n* es 0? No puede restarle 1 (los registros no pueden tener como contenido enteros negativos: no puedes sacar un frijol de una caja vacía), así que, obstaculizada, tiene que *ramificarse*. Es decir, tiene que ir a algún otro lugar del programa para recibir su siguiente instrucción. Para esto se necesita que todas las instrucciones de *disminución* anoten el lugar del programa al que debe irse si el contenido del registro actual es 0. Así, la definición completa de *disminución* es:

Registra una *disminución* a *n* (resta 1 al contenido del registro *n*) si puedes y ve al paso *m, o bien,* si no puedes registrar una disminución a *n, ramifica* al paso *p*.

He aquí, pues, la lista completa de todo lo que una máquina de registro puede hacer, con prácticos nombres abreviados: *Fin, Inc* y *Dir* (por *di*sminuye o *r*amifica).

Puede ser que a primera vista no te parezca que una máquina así de simple pudiera hacer algo muy interesante que digamos: todo lo que puede hacer es *poner un frijol en la caja* o *sacar un frijol de la caja* (si puede encontrarlo, y ramificarse a otra instrucción si no). Sin embargo, lo cierto es que puede computar cualquier cosa que una computadora pueda computar.

Empecemos con una simple adición. Supón que quisieras que la máquina de registro *sumara* el contenido de un registro (digamos, el registro 1) al contenido de otro registro (registro 2). Así, si el registro 1 tiene contenido [3] y el registro 2 tiene contenido [4], queremos que el programa acabe con que el registro 2 tenga contenido [7], pues 3 + 4 = 7. He aquí un programa que podría hacerlo, escrito en un lenguaje simple que podemos llamar PER, por Programación Ensambladora de Registros:

Programa 1: suma [1,2]

Paso	*Instrucción*	*Registro*	*Ve al paso*	*[Ramifica al paso]*
1	*Dir*	1	2	3
2	*Inc*	2	1	
3	*Fin*			

Las primeras dos instrucciones forman un simple bucle, que disminuye el registro 1 e

incrementa el registro 2, una y otra vez, *hasta que el registro 1 esté vacío,* lo que la unidad procesadora «advierte», y enseguida se *ramifica* al paso 3, el cual le ordena detenerse. La unidad procesadora no puede saber cuál es el contenido de un registro salvo cuando es 0. En cuanto a la imagen de los frijoles en la caja, puedes pensar que la unidad procesadora es ciega, incapaz de ver qué hay en un registro hasta que esté vacío, algo que puede detectar a tientas. Sin embargo, a pesar del hecho de que por lo general no puede saber cuál es el contenido de sus registros, cuando se ejecuta el programa 1 *siempre* sumará el contenido del registro 1 (sin importar qué número haya en el registro 1) al contenido del registro 2 y luego se detendrá. (¿Te das cuenta de por qué esto casi siempre funciona? Repasa algunos ejemplos para asegurarte.) He aquí una manera llamativa de verlo: la máquina de registro puede perfectamente sumar dos números sin saber qué números son (¡o qué cosa son los números o qué significa sumar!).

Ejercicio 1

a) ¿Cuántos pasos le tomará a la máquina de registro sumar 2 + 5 y obtener 7 ejecutando el programa 1 (contando fin *como paso)?*

b) ¿Cuántos pasos le tomará sumar 5 + 2? (¿Qué conclusión sacas de esto?)[4]

Este proceso puede diagramarse en lo que se conoce como un *diagrama de flujo.* Cada círculo representa una instrucción. El número dentro del círculo representa la *dirección* del registro que va a manipularse (no el contenido de un registro), «+» representa *Inc* y «–» representa *Dir.* El programa siempre comienza en α, alfa, y se detiene cuando llega a Ω, omega. Las flechas llevan a la siguiente instrucción. Obsérvese que toda instrucción *Dir* tiene dos flechas que salen, una que indica adónde ir cuando puede disminuir y una que indica adónde ir cuando no puede porque el contenido del registro es 0 (o *ramifica en cero*).

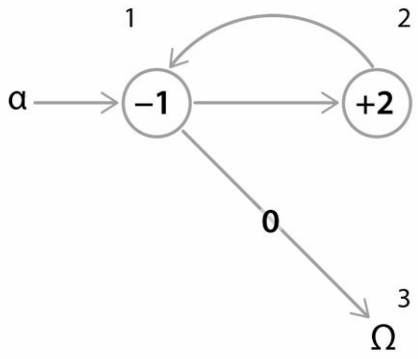


FIGURA 2

Ahora escribamos un programa que simplemente *mueva* el contenido de un registro a

otro:

Programa 2: mueve [4,5]

Paso	*Instrucción*	*Registro*	*Ve al paso*	*[Ramifica al paso]*
1	*Dir*	5	1	2
2	*Dir*	4	3	4
3	*Inc*	5	2	
4	*Fin*			

Éste es el diagrama de flujo:

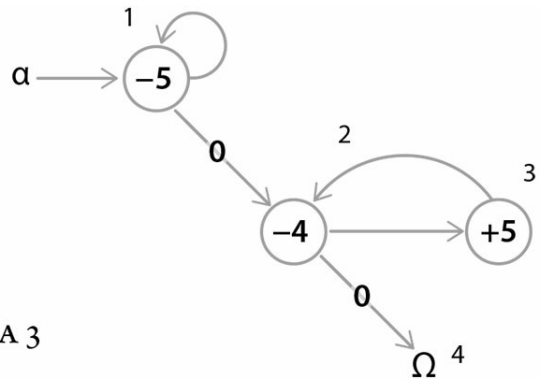


FIGURA 3

Observa que el primer bucle de este programa limpia el registro 5, de manera que fuera cual fuera su contenido en un inicio, éste no contaminará lo que se consigue en el registro 5 junto al segundo bucle (que es precisamente nuestro bucle de suma, que suma el contenido del registro 4 al 0 del registro 5). Este paso para inicializar se conoce como reducir a cero el registro, y es una operación estándar muy útil. Con ella constantemente prepararás registros para usarse.

Un tercer programa simple *copia* el contenido de un registro en otro, sin cambiar en nada el contenido original. Estudia el diagrama de flujo y luego el programa:

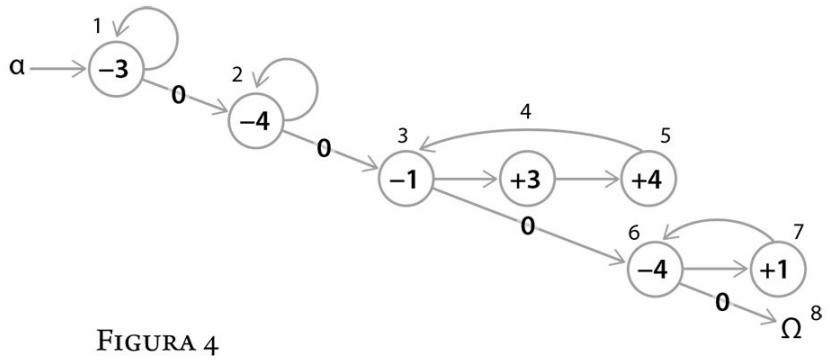


FIGURA 4

Programa 3: copia [1,3]

Paso	*Instrucción*	*Registro*	*Ve al paso*	*[Ramifica al paso]*
1	*Dir*	3	1	2
2	*Dir*	4	2	3
3	*Dir*	1	4	6
4	*Inc*	3	5	
5	*Inc*	4	3	
6	*Dir*	4	7	8
7	*Inc*	1	6	
8	*Fin*			

No cabe duda de que ésta es una manera enredada de copiar, pues lo hacemos primero *moviendo* el contenido del registro 1 al registro 3 mientras hacemos una copia en el registro 4, y luego movemos esa copia de vuelta al registro 1. Sin embargo, funciona, siempre funciona. Sin importar cuál sea el contenido de los registros 1, 3 y 4 en un inicio, cuando el programa se detiene, lo que haya estado en el registro 1 seguirá allí y en el registro 3 habrá una copia de ese contenido.

Si el funcionamiento de este programa aún no te parece completamente obvio, saca unas tazas para los registros (en cada taza anota un número, o sea su dirección) y un montón de monedas de cincuenta centavos (o frijoles), y «simula a mano» todo el proceso. Pon *unas cuantas* monedas en cada registro y toma nota de cuántas pones en el registro 2 y en el registro 3. Si sigues el programa al pie de la letra, cuando termines, la cantidad de monedas en el registro 1 será la misma que al principio, y ahora habrá la misma cantidad en el registro 3. *Es muy importante que internalices los procesos básicos de la máquina de registro,* hasta el punto de no tener que pensar arduamente en ellos, porque en las páginas siguientes vamos a explotar este nuevo talento. Así, pues, date unos minutos para *convertirte* en una máquina de registro (tal como un actor puede

convertirse en Hamlet).

He visto que algunos de mis alumnos cometen un error muy simple: imaginan que cuando decrementan un registro tienen que poner esa moneda, la que acaban de quitar del registro *n,* en algún otro registro. No. Las monedas decrementadas van de vuelta al montón, tu «infinita» dotación de monedas para esta simple rutina de sumar y restar.

Con *mover, copiar* y *reducir a cero* entre nuestras herramientas, estamos listos para regresar a nuestro programa de adición y mejorarlo. El programa 1 pone la respuesta correcta a nuestro problema de adición en el registro 2, pero en el proceso destruye los contenidos originales de los registros 1 y 2. Posiblemente nos gustaría tener un programa de adición más sofisticado que salvara estos valores para algún uso posterior y que pusiera la respuesta en algún otro lugar. Consideremos, pues, la tarea de sumar el contenido del registro 1 al contenido del registro 2, poniendo la respuesta en el registro 3 y dejando intactos los contenidos de los registros 1 y 2.

Este diagrama de flujo lo consigue:

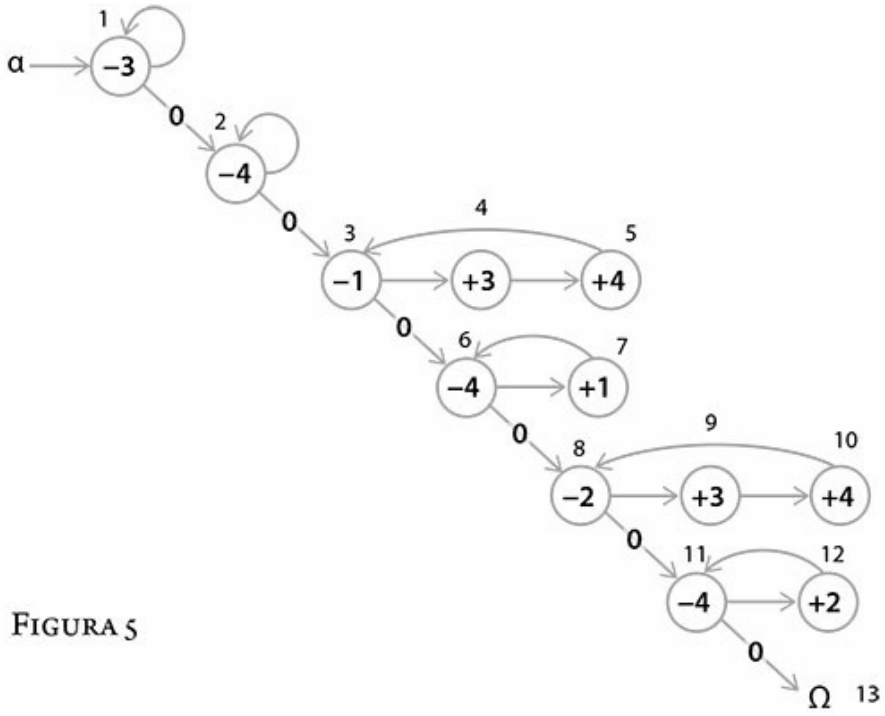


FIGURA 5

Podemos analizar los bucles para ver qué hace cada uno. Primero reducimos a cero el *registro de respuesta,* el registro 3, y luego reducimos a cero un registro libre (el registro 4) para usarlo como un tinaco temporal o *memoria intermedia.* Luego copiamos el contenido del registro 1 en los registros 3 y 4, y devolvemos ese contenido de la memoria intermedia al registro 1, y así lo restablecemos (y en el proceso reducimos a cero el registro 4 para volverlo a usar como memoria intermedia). Luego repetimos esta operación usando el registro 2, con el efecto de que sumamos el contenido del registro 2

al contenido que ya movimos al registro 3. Cuando el programa se detiene, la memoria intermedia 4 está de nuevo vacía, la respuesta está en el registro 3, y los dos números que sumamos están de regreso en su sitio original, los registros 1 y 2.

El programa PER de 13 pasos (página siguiente) pone toda la información del diagrama de flujo en la forma que la unidad de procesamiento puede leer.

No voy a recomendar que simules este programa con las copas y las monedas. La vida es breve, y *una vez que has internalizado los procesos básicos en tu imaginación,* ahora puedes aprovechar las ventajas de un aparato ortopédico: RodRego, una máquina de registro que puedes descargar de ‹http://sites.tufts.edu/rodrego/›.

Hay versiones de RodRego para Mac y para PC disponibles para ejecutarse en tu computadora. Creamos esta herramienta de pensamiento hace más de 20 años en el

Programa 4: suma [1,2,3] no destructiva

Paso	*Instrucción*	*Registro*	*Ve al paso*	*[Ramifica al paso]*
1	*Dir*	3	1	2
2	*Dir*	4	2	3
3	*Dir*	1	4	6
4	*Inc*	3	5	
5	*Inc*	4	3	
6	*Dir*	4	7	8
7	*Inc*	1	6	
8	*Dir*	2	9	11
9	*Inc*	3	10	
10	*Inc*	4	8	
11	*Dir*	4	12	13
12	*Inc*	2	11	
13	*Fin*			

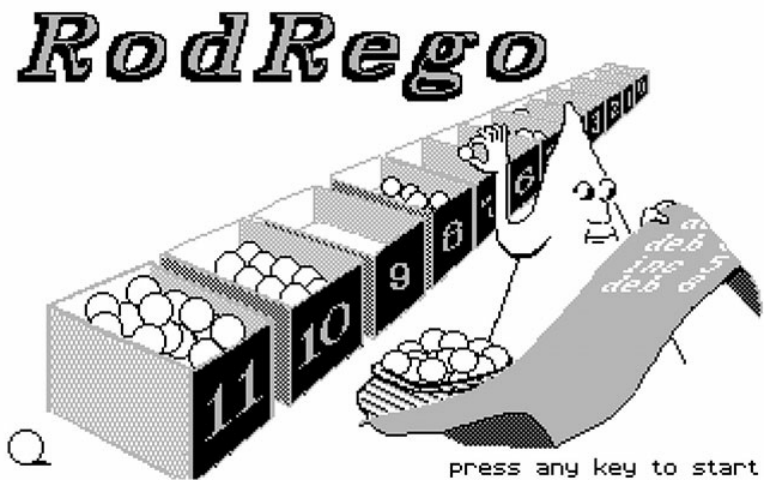


Figura 6. *Pantalla de inicio de la máquina de registro RodRego original, 1986*

Estudio de Software Curricular, y cientos de estudiantes y otras personas la han usado para ser muy desenvueltos al pensar como máquina de registro. Allí puedes teclear tus programas PER y verlos ejecutarse, ya sea con frijoles o números en los registros. También hay animaciones de PowerPoint que muestran el camino que toma la unidad procesadora a lo largo del diagrama de flujo para la suma, por ejemplo, así que puedes ver la correspondencia exacta de las instrucciones PER con los círculos del diagrama de flujo.

Ahora pasemos a la resta. Aquí hay un primer intento de diagrama de flujo para restar el contenido del registro 2 del contenido del registro 1, y poner la respuesta en el registro 4. ¿Notas qué es lo que falla?

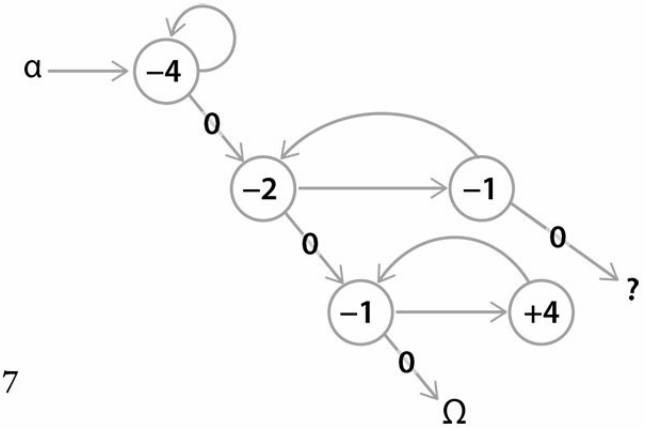


FIGURA 7

Esto funcionará únicamente cuando el contenido del registro 1 sea mayor que el contenido del registro 2. ¿Y qué cuando no es así? El registro 1 se reducirá a cero avanzando hasta la mitad de un paso en el bucle de resta antes de poder terminar la

operación. ¿Qué debe pasar entonces? No podemos simplemente pedirle a la computadora que se detenga, pues esto deja una respuesta equivocada (0) en el registro 4. Podemos usar esta reducción a cero para empezar un nuevo proceso, que primero retrocede medio bucle y deshace la disminución provisional del registro 2. En este punto, el contenido del registro 2 (no del registro 1) da la respuesta correcta si lo interpretamos como un número *negativo,* así que puedes simplemente *mover* ese contenido al registro 4 (que ya está reducido a cero) y en algún lugar poner un signo que indique que la respuesta es un número negativo. La acción más obvia es reservar un registro precisamente para esta tarea: digamos que el registro 3. Redúcelo a cero al principio, junto con el registro 4, y luego haz que el programa ponga una «bandera» en el registro 3 como signo de la respuesta, donde 0 significa + y 1 significa −. Presento a continuación el diagrama de flujo con comentarios que explican qué hace cada paso o bucle (puedes poner este tipo de comentarios en tus programas PER entre marcas #; son para ti y para otros seres humanos: RodRego no les hará caso).

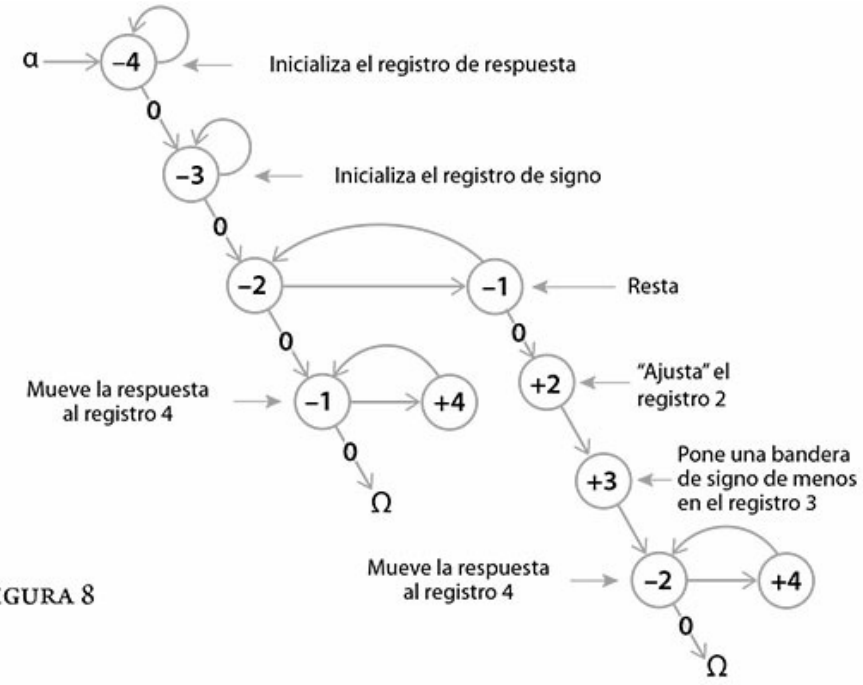


FIGURA 8

Ejercicio 2

a) Escribe el programa PER para este diagrama de flujo. (Observa que, como el programa se ramifica, puedes numerar los pasos de varias maneras diferentes. No importa cuál escojas siempre y cuando los comandos «ve *a» apunten a los pasos correctos.)*

b) ¿Qué pasa cuando el programa trata de restar 3 de 3 o 4 de 4?

c) ¿Qué error posible se evita reduciendo a cero el registro 3 antes de tratar

de restar en el paso 3 y no después del paso 4?

Con la suma y la resta en nuestro haber, es fácil concebir la multiplicación y la división. Multiplicar *n* por *m* no es más que sumar *n* a cero *m* número de veces. Así, podemos ordenarle a la computadora que haga exactamente eso, usando un registro como *contador,* que haga una cuenta regresiva de *m* a 0 decrementando una vez cada que se complete el bucle de suma.

Ejercicio 3

a) Dibuja un diagrama de flujo (y escribe el programa PER*) para multiplicar el contenido del registro 1 por el contenido del registro 3, y pon la respuesta en el registro 5.*

b) (Opcional.)[5] *Copiando y moviendo, mejora el multiplicador que creaste en el problema a: cuando se detiene, los contenidos originales de los registros 1 y 3 se restablecen, así que fácilmente puedes revisar que las entradas y salidas sean las correctas después de una ejecución.*

c) (Opcional.) Dibuja un diagrama de flujo y escribe un programa PER *que examine los contenidos de los registros 1 y 3 (¡sin destruirlos!) y escriba la dirección (1 o 3) del contenido del registro 2, que es mayor, y ponga 2 en el registro 2 si los contenidos de los registros 1 y 3 son iguales (después de que este programa se haya ejecutado, los contenidos de los registros 1 y 3 no deberían haber cambiado, y el registro 2 debería decir si sus contenidos son iguales y, si no, cuál de ellos tiene el contenido que es mayor).*

Del mismo modo, la división puede hacerse restando el divisor del dividendo una y otra vez y contando la cantidad de veces que lo hacemos. Podemos dejar el residuo, si lo hay, en un registro especial de residuo. No obstante, aquí debemos tener la precaución de añadir una medida de seguridad: no debemos dividir entre cero (¿o sí?), así que antes de que comience cualquier división, debemos ejecutar una simple revisión del divisor, tratando de decrementarlo. Si podemos disminuirlo, deberíamos incrementarlo sólo una vez (y restablecer su valor adecuado) y luego proceder con la división. Sin embargo, si llegamos a cero al tratar de disminuirlo, debemos dar la alarma. Podemos hacerlo reservando un registro para una bandera de ERROR: un 1 en el registro 5 puede significar «¡PELIGRO! ¡Me acaban de pedir que divida entre cero!»

Aquí está el diagrama de flujo para la operación de dividir el contenido del registro 1 entre el contenido del registro 2 y poner la respuesta en el registro 3, el residuo en el registro 4 y resaltar el registro 5 con un «mensaje de error» (un 1 significa «Me pidieron que dividiera entre cero»).

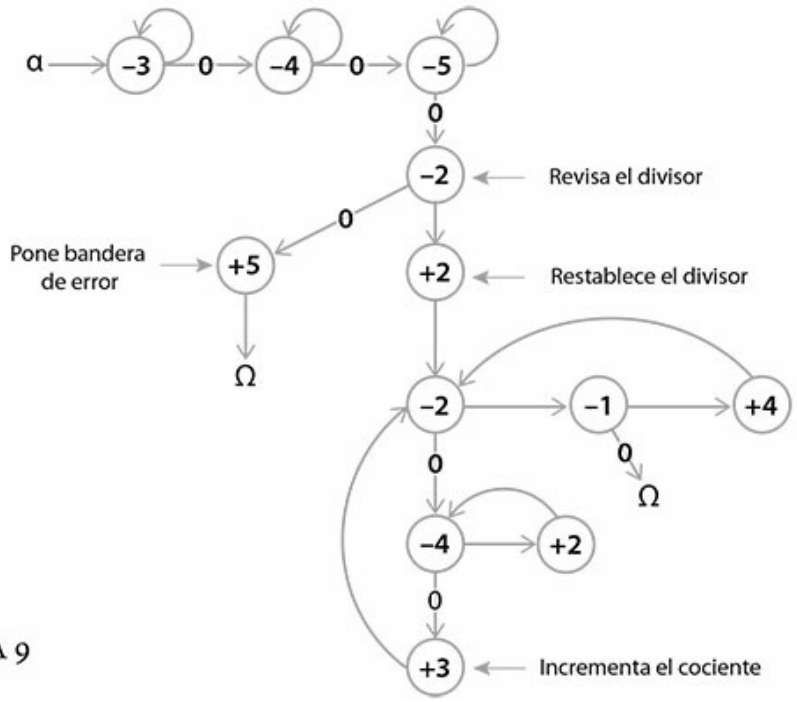


FIGURA 9

Repasa el diagrama de flujo y observa cómo el cero en el divisor aborta la operación y levanta una bandera. Observa también que el registro 4 tiene una doble obligación: no sólo sirve como copia del divisor y restablece el divisor para toda resta subsecuente, sino también como un registro de residuos potencial. Si el registro 1 se reduce a cero antes de que el registro 4 pueda regresar su contenido al registro 2 para otra resta, ese contenido es el residuo, y está justo donde le corresponde.

SECRETO 1. Competencia sin comprensión: algo —*e. g.,* una máquina de registro— puede hacer una perfecta aritmética sin tener que comprender lo que hace.

La máquina de registro no es una mente. No comprende nada. Sin embargo, maomeno comprende tres cosas simples —*Inc, Dir* y *Fin*— en el sentido de que obedientemente lleva a cabo estas tres «instrucciones» cada vez que se le presentan. No son instrucciones *reales,* desde luego; son maomeno instrucciones. Nosotros las vemos como instrucciones, y la máquina de registro las lleva a cabo como si lo fueran, así que es de lo más práctico llamarlas instrucciones.

Como ahora puedes ver, *Dir,* disminuye o ramifica, es la clave para el poder de la máquina de registro. Es la única instrucción que le permite a la computadora «detectar» (maomeno detectar) cualquier cosa del mundo y usar lo que detecta para guiar su siguiente paso. Y de hecho esta *ramificación condicional* es la clave del poder de todas las computadoras que almacenan programas, un hecho que Ada Lovelace reconoció ya en el siglo XIX cuando escribió su brillante discusión del Motor Analítico de Charles

Babbage, el prototipo de todas las computadoras.[6]

Ensamblar estos programas a partir de sus partes puede volverse un ejercicio algo rutinario ya que lo empezamos a dominar. De hecho, una vez que hayamos compuesto cada una de las rutinas aritméticas, podremos usarlas una y otra vez. Supón que las numeramos, de modo que SUMA fuera la operación 0, RESTA fuera la operación 1, MULTIPLICA fuera la operación 2, etc. COPIA sería la operación 5, MUEVE sería la operación 6, etc. Entonces podríamos usar un registro para almacenar una instrucción numerada.

Ejercicio 4 (opcional)
Dibuja un diagrama de flujo y escribe un programa PER que convierta una máquina de registro en una simple calculadora de bolsillo, así:

a) Usa el registro 2 para la operación:
0 = SUMA
1 = RESTA
2 = MULTIPLICA
3 = DIVIDE
b) Pon los valores que deben seguirse en los registros 1 y 3.
(Por lo tanto, 3 0 6 significaría 3 + 6, y 5 1 3 significaría 5 – 3, y 4 2 5 significaría 4 × 5, y 9 3 3 significaría 9 ÷ 3). Luego pon los resultados de la operación en los registros 4 a 7 y usa el registro 4 para el signo (con 0 para + y 1 para –) y el registro 5 para la respuesta numérica, el registro 6 para cualquier residuo en el caso de una división, y el registro 7 como alarma que señale un error en la entrada (ya sea «divide entre cero» o una operación sin definir en el registro 2).

Observa que en este ejemplo estamos usando el contenido de los registros (en cada caso un número) para simbolizar cuatro cosas muy diferentes: un número, una operación aritmética, el signo de un número y una bandera de error.

SECRETO 2. Qué simbolice el número de un registro depende del programa que hayamos creado.

Con los bloques que ya hemos creado podemos construir operaciones más imponentes. Con la suficiente paciencia podríamos dibujar el diagrama de flujo y escribir el programa para que ELEVE AL CUADRADO el número del registro 7, o un programa para que CALCULE EL PROMEDIO de los contenidos de los registros 1 a 20, o ENCUENTRE LOS FACTORES del contenido del registro 6 y ponga un 1 en el registro 5 si 5 es uno de los factores, o compare los contenidos de los registros 3 y 4 y ponga el contenido mayor en

el registro 5 a menos que sea exactamente el doble, y en tal caso ponga una bandera en el registro 7. Y así sucesivamente.

Una rutina especialmente útil sería BUSCAR entre 100 registros para ver si alguno de ellos tiene determinado contenido y poner el nombre de la *dirección* de ese registro en el registro 101. (¿Cómo hacer esto? Pon el número OBJETIVO en el registro 102, y una copia del objetivo en el registro 103; reduce a cero el registro 101, y luego, empezando en el registro 1, resta su contenido del de 103 [después de incrementar el registro 101], en busca de una respuesta de cero. Si no la obtienes, avanza al registro 2, y así sucesivamente. Si algún registro tiene el número objetivo, detente; la dirección de ese registro estará en el registro 101.) Gracias al poder «sensorial» básico plasmado en *Dir* —su capacidad de «detectar» un cero cuando trata de disminuir un registro— podemos hacer que la máquina de registro vuelva sus «ojos» a sí misma para que pueda examinar sus propios registros y mueva contenidos de un lado a otro y cambie de operación según qué encuentre y dónde.

SECRETO 3. Como un número en un registro puede representar lo que sea, esto significa que la máquina de registro puede, en principio, diseñarse para «detectar» cualquier cosa, «discriminar» cualquier rasgo repetido o característica que pueda asociarse con un número —o una cantidad de números—.

Por ejemplo, una fotografía en blanco y negro —*cualquier* fotografía en blanco y negro, o incluso la foto de esta página— puede representarse con un gran banco de registros, con un registro por cada pixel, con 0 para un punto blanco y 1 para un punto negro. Ahora escribe el programa para la máquina de registro que pueda revisar miles de fotografías en busca de una foto de una línea recta negra horizontal sobre un fondo blanco. (Mejor lo no intentes. La vida es breve. Tan sólo imagina con algún detalle el difícil y enormemente tardado proceso con el que podría hacerse esto.) Ya que hayas creado —en tu imaginación— tu reconocedor de líneas horizontales, tu reconocedor de líneas verticales y tu reconocedor de semicírculos, piensa cómo podrías unirlos con unos cuantos prácticos discernidores más (o unas cuantas decenas) y hacer algo que pudiera discernir una letra «A» (mayúscula) —¡en cientos de fuentes distintas!—. Éste es uno de los triunfos más o menos recientes de la programación, el *software* de Reconocimiento Óptico de Caracteres u OCR, por las siglas de *Optical Character Recognition,* que puede escanear una página impresa y convertirla de manera bastante fiable en un archivo informático de texto (en el que cada signo alfabético o numérico es representado por un número en código ASCII, de modo que puede hacerse búsqueda de texto, así como el resto de la magia del procesamiento de palabras… sin nada más que aritmética). ¿Un programa OCR puede *leer*? En realidad no; no entiende lo que se le pone enfrente.

Maomeno lee, que es una competencia maravillosamente útil que puede sumarse a nuestro generoso estuche de partes móviles.

SECRETO 4. Como un número puede representar lo que sea, puede representar una instrucción o una dirección.

Podemos usar un número en un registro para que represente una instrucción, como SUMA, RESTA, MUEVE o BUSCA, y para que represente direcciones (registros en la computadora), de modo que podemos almacenar toda una secuencia de instrucciones en una serie de registros. Si entonces tenemos un programa central (programa A) que le ordene a la máquina que vaya de registro en registro haciendo lo que cada uno de ellos le ordene, entonces podemos almacenar en esos registros un segundo programa B. Cuando encendemos la máquina y hacemos que ejecute el programa A, lo primero que hace es consultar los registros que le dicen que ejecute el programa B, lo cual hace a continuación. Esto significa que podríamos almacenar el programa A de una vez por todas en la unidad de procesamiento central de la máquina de registro, en un conjunto reservado de registros (podría ser «f*irmware*» quemado en la memoria ROM, de sólo lectura), y luego usar el programa A para ejecutar los programas B, C, D, etc., según qué números pongamos en los registros regulares. Al instalar el programa A en nuestra máquina de registro la convertimos en una *computadora de programa almacenado.*

El programa A le da a nuestra máquina de registro la competencia de ejecutar fielmente cualesquiera instrucciones que pongamos (por número) en sus registros. Todo programa posible que sea capaz de ejecutar consiste en una serie de números, ordenados, que el programa A consultará, en orden, y hará cualquier cosa que cada número especifique. Y si creamos un sistema para expresar estas instrucciones de una manera no ambigua (por ejemplo, poner el requisito de que los nombres de las instrucciones sean todos del mismo tamaño, digamos dos dígitos) podemos tratar toda la serie de números que constituyen el programa B, digamos

86, 92, 84, 29, 08, 50, 28, 54, 90, 28, 54, 90

como un solo grande y largo número:

869284290850285490285490.

Este número es a la vez el «nombre» exclusivo del programa —el programa B— y el programa mismo, que el programa A ejecuta paso por paso. Otro programa es

284570297590287529075489274902754248509284285404 23,

y otro es

8908296472490284952498856743390438503882459802854544254789653985,

pero la mayoría de los programas interesantes tendrían nombres muchísimo más largos, del tamaño de millones de dígitos. Los programas que tienes almacenados en tu computadora portátil, como un procesador de palabras y un navegador, son esos números largos, del tamaño de muchos millones de dígitos (binarios). Un programa del tamaño de 10 megabytes tiene una cadena de 80 millones de ceros y unos.

SECRETO 5. A todos los programas posibles puede ponérseles como nombre un número exclusivo, que luego puede tratarse como una lista de instrucciones para que las ejecute una máquina universal.

Alan Turing fue el brillante teórico y filósofo que ideó este esquema con otra sencilla computadora imaginaria, una que avanza a trompicones de un lado a otro sobre una cinta de papel dividida en cuadros, e hizo que su comportamiento dependiera (¡*ajá!*, ramificación condicional) de si lee un cero o un uno en el cuadro que en ese momento está bajo su cabeza lectora. Todo lo que la máquina de Turing puede hacer es voltear el bit (borrar 0 y escribir 1, o viceversa) o dejarlo en paz, y luego mover a la izquierda o a la derecha un cuadro de cinta *y avanzar a su siguiente instrucción.* Creo que el lector estará de acuerdo conmigo en que escribir programas de la máquina de Turing para que SUMEN, RESTEN y realicen otras funciones, usando únicamente los números binarios 0 y 1 en vez de todos los números naturales (0, 1, 2, 3, 4, 5, etc.) y moviendo un cuadro cada vez es un ejercicio más descomunal que los de nuestra máquina de registro, pero lo que Turing proponía era exactamente lo mismo. Una máquina de Turing universal es un dispositivo con un programa A (integrado si tú quieres) que le permite «leer» en la cinta su programa B y luego ejecutarlo usando cualquier otra cosa que haya en la cinta como datos o entrada del programa B. La máquina de registro de Hao Wang puede ejecutar cualquier programa que sea posible reducir a aritmética y ramificación condicional, igual que la máquina de Turing. Ambas máquinas tienen el maravilloso poder de tomar el *número* de cualquier otro programa y ejecutar *ese* programa. En lugar de fabricar miles de máquinas de cómputo diferentes, cada una configurada para ejecutar determinada tarea complicada, fabricamos una sola máquina universal (con el programa A instalado), y luego podemos hacer que cumpla nuestras órdenes alimentándola de programas —*software*— que crean **máquinas virtuales**.

La máquina universal de Turing es, en otras palabras, una imitadora universal. Igual

nuestra menos famosa máquina universal de registro. Igual tu computadora portátil. No hay nada que pueda hacer tu computadora portátil que no pueda hacer la máquina universal de registro, y viceversa. Pero mejor espera sentado: nadie dijo que todas las máquinas tuvieran la misma velocidad. Ya hemos visto que nuestra máquina de registro es dolorosamente lenta en algo tan laborioso como la división, que, por Dios santo, hace mediante restas seriales. ¿No hay manera de acelerar las cosas? Sí que la hay. De hecho, la historia de las computadoras desde la época de Turing es justamente la historia de las maneras cada vez más rápidas de hacer lo mismo que la máquina de registro, y nada más.

SECRETO 6. Todas las mejoras a las computadoras desde que Turing inventó su imaginaria máquina de cintas de papel son simples maneras de hacerlas más veloces.

Por ejemplo, John von Neumann creó la arquitectura para la primera computadora funcional seria, y a fin de acelerarla, amplió la ventana o cabeza lectora de la máquina de Turing de un bit cada vez a muchos bits cada vez. Muchas de las primeras computadoras leían «palabras» de ocho o 16 bits, e incluso palabras de 12 bits. Hoy en día, las palabras de 32 bits son ampliamente utilizadas. Esto sigue siendo un cuello de botella —el cuello de botella de Von Neumann—, ¡pero es 32 veces más amplio que el cuello de botella de la máquina de Turing! Simplificando un poco, podemos decir que todas las palabras se COPIAN de la memoria una por una en un registro especial (el registro de instrucciones), donde se LEEN y ejecutan. Normalmente una palabra tiene dos partes, el código de operación (*e. g.,* SUMA, MULTIPLICA, MUEVE, COMPARA, SI ES CERO SALTA) y una dirección, que le dice a la computadora a qué registro ir por el contenido sobre el que hay que operar. Así, **10101110** 11101010101 podría decirle a la computadora que realice la operación **10101110** en el contenido del registro 11101010101, y ponga la respuesta, como siempre, en un registro especial llamado el acumulador. La gran diferencia entre la máquina de registro y una máquina de Von Neumann es que la primera puede operar sobre cualquier registro (sólo *Inc* y *Dir,* desde luego), mientras que una máquina de Von Neumann hace todo el trabajo aritmético en el acumulador, y simplemente COPIA y MUEVE (o ALMACENA) contenidos a los registros que conforman la memoria. Compensa todos estos movimientos y copiados adicionales el hecho de que sea capaz de realizar muchas operaciones fundamentales diferentes, integradas todas ellas. Esto es, hay un circuito electrónico especial para SUMA y otro para RESTA, y todavía uno más para SI ES CERO SALTA, y así sucesivamente. El código de operación es algo así como un código de área en el sistema telefónico o un código postal en el correo: manda aquello con lo que esté trabajando al lugar indicado para que se ejecute. Así es como el *software* se encuentra con el *hardware*.

¿Cuántas operaciones primitivas se ejecutan en las computadoras reales en estos tiempos? Puede haber cientos, miles o, en un retorno a los viejos tiempos, una computadora puede ser una computadora de conjunto de instrucciones reducidas (RISC, *reduced instruction set computer*), que se las arregla con unas cuantas decenas de operaciones primitivas pero lo compensa con la deslumbradora velocidad con que las ejecuta. (Si las instrucciones *Inc* y *Dir* pudieran realizarse un millón de veces más rápido que una operación integrada de SUMA, convendría componer SUMA usando *Inc* y *Dir,* como hicimos antes, y saldríamos ganando con todas las sumas de menos de un millón de pasos.)

¿Cuántos registros hay en las computadoras actuales? Millones, o incluso miles de millones (pero cada uno es finito, así que los números verdaderamente grandes tienen que extenderse a lo largo de grandes cantidades de registros). Un byte tiene 8 bits. Si en tu computadora tienes 64 megabytes de memoria RAM (por *random access memory,* memoria de acceso aleatorio), tienes 16 millones de registros de 32 bits, o el equivalente. Vimos ya que los números en los registros pueden representar otras cosas que no sean enteros positivos. Los números *reales* (como π, $\sqrt{2}$ o 1/3) se almacenan usando un sistema de representaciones de «coma flotante», que rompe el número en dos partes, la *base* y el *exponente,* como en la notación científica (1.495×10^{41}), lo que le permite a la aritmética de la computadora manejar (aproximaciones de) números que no sean los naturales. Las operaciones de coma flotante son simplemente operaciones aritméticas (en particular multiplicaciones y divisiones) que usan como valores estos números de coma flotante, y la supercomputadora más rápida que podías comprar hace 20 años (cuando escribí la primera versión de este apartado) podía realizar más de 4 MEGAFLOPS: más de 4 millones de operaciones de coma flotante por segundo (del inglés *flops: fl*oating point *o*perations *p*er *s*econd).

Si eso no es suficientemente rápido para ti, puede ayudarte unir una cantidad de esas máquinas en paralelo, para que todas trabajen al mismo tiempo, y no en serie, haciendo cola en espera de resultados con los cuales trabajar. No hay nada que esa máquina paralela pueda hacer que una meramente serial no pueda hacer, más despacio. De hecho, la mayoría de las máquinas paralelas que se han estudiado activamente en los últimos 20 años han sido simulaciones de **máquinas virtuales** en máquinas Von Neumann estándar (no paralelas). Ya se ha creado *hardware* paralelo para propósitos especiales, y los diseñadores de computadoras están muy ocupados explorando los costos y beneficios que tendría ampliar el cuello de botella de Von Neumann, y acelerar el tráfico que pasa a través de él, de todas las maneras posibles: con coprocesadores, memorias caché y varios otros accesos. Hoy en día, la computadora japonesa K de Fujitsu puede operar a 10.51 PETAFLOPS, es decir, más de 10 000 *billones* de operaciones de coma flotante por segundo.

Eso *podría* ser casi lo suficientemente rápido para simular la actividad informática de tu cerebro conforme éste funciona. Tu cerebro es un procesador paralelo por excelencia, con cerca de 100 000 millones de neuronas, y cada una de ellas es un pequeño agente bastante complicado que tiene sus propias prioridades. El «nervio» óptico, que lleva información visual de tu ojo a tu cerebro, tiene, por sí solo, varios millones de canales (neuronas) de ancho. Sin embargo, las neuronas trabajan muchísimo más despacio que los circuitos de las computadoras. Una neurona puede cambiar de estado y mandar una pulsación (es verosímil que sea su versión de *Inc* o *Dir*) en unos cuantos milisegundos —milésimas, no millonésimas ni milmillonésimas de segundo—. Las computadoras mueven bits de un lado a otro a una velocidad cercana a la de la luz: por esa razón, hacer computadoras más chicas es una movida clave para hacerlas más veloces. A la luz le toma aproximadamente una milmillonésima de segundo viajar 30 centímetros, así que si quieres que dos procesos se comuniquen más rápido que eso, tienen que estar más cerca uno de otro.

SECRETO 7. ¡No hay más secretos!

Tal vez la característica más maravillosa de las computadoras sea que como están construidas con pasos sencillos y hechas de partes (operaciones) que también son de lo más simple, en ellas sencillamente no hay espacio para que tengan ningún secreto bajo la manga. Nada de ectoplasma, «resonancias mórficas», campos de fuerza invisibles, leyes físicas desconocidas hasta hoy, nada de tejido maravilloso. *Sabes* que si consigues que un programa de computadora modele algún fenómeno, en el modelo no intervienen más causas que las que están compuestas de todas las operaciones aritméticas.

¿Y qué decir de la informática cuántica, que causa furor en estos días? ¿No pueden las computadoras cuánticas hacer cosas que ninguna computadora ordinaria puede? Sí y no. Lo que pueden hacer es resolver muchos problemas, computar muchos valores simultáneamente, gracias a la «superposición cuántica», la extraña y delicada propiedad en la que una entidad no observada puede estar en «todos los estados posibles» al mismo tiempo, hasta que la observación provoca el «colapso del paquete de ondas» (para más sobre el tema, consulta tu libro o sitio web favorito de física para todos). Básicamente, una computadora cuántica es tan sólo la última —y muy impresionante— innovación en velocidad: un salto cuántico, podría decirse, en velocidad de procesamiento. Una máquina de Turing que avanza a trompicones sobre su cinta de papel, o una máquina registradora que va de un lado a otro incrementando y disminuyendo registros únicos, tiene límites muy estrictos sobre lo que puede hacer en pedazos de tiempo que en la práctica son pequeños: minutos, horas o días. Una supercomputadora como la K de Fujitsu puede hacer las mismas cosas billones de veces más rápido, pero eso todavía no

es suficiente para resolver algunos problemas, especialmente unos de criptografía. Es ahí donde la velocidad extra de las computadoras cuánticas podría redituar: si la gente puede resolver los endemoniadamente difíciles problemas de ingeniería que se presentan al tratar de hacer una computadora cuántica práctica y estable. Podría no ser posible, en cuyo caso tal vez tengamos que conformarnos con apenas algunos trillones de FLOPS.

25. Máquinas virtuales

Las máquinas reales están hechas de partes móviles materiales y normalmente se les dan nombres acordes con los trabajos para los que fueron concebidas. Las podadoras, los abrelatas y los molinillos de café tienen variedad de diseños, que en ocasiones trabajan de acuerdo con diferentes principios físicos, pero tienen en común que las máquinas con un mismo nombre *hacen el mismo trabajo* en algún nivel de descripción. A lo mejor algunas lo hacen mejor, pero todo depende de lo que quiera el consumidor. El propietario de una casa con jardín puede preferir una podadora lenta pero silenciosa; el dueño de una cafetería puede preferir un molinillo al que se le puede ajustar con más exactitud el tipo de molido a costa de que sea más difícil de operar. Algunas máquinas son versátiles: si enchufas un accesorio, puedes convertir un taladro en una sierra o una lijadora. Así son las computadoras, sólo que en lugar de que puedan hacer decenas de cosas distintas, pueden hacer archimillones. Y en lugar de que tenga que enchufarse un accesorio diferente para cada tarea, se abre un programa diferente —una muy larga cadena de ceros y unos— que cambia todos los interruptores internos necesarios a los ajustes correctos para cumplir con la tarea. Cada uno de estos diferentes sistemas de ajustes es una máquina diferente: una máquina *virtual* diferente, una máquina «hecha de instrucciones», no de engranajes y rodamientos, cables y poleas. En las computadoras, las instrucciones pueden tomar el lugar de los engranajes y rodamientos porque en lugar de procesar masa para pan, pulpa de papel o lingotes de hierro, las computadoras procesan información, y la información siempre puede traducirse a código binario, cero y uno, el único que la computadora puede —o necesita— «leer». Billones de esos ceros y unos van y vienen por los circuitos impresos en los chips de silicio; abren y cierran puertas temporalmente, mueven los flujos de información a uno u otro circuito y así regulan lo que pasa con ella. Los millones de pequeños lugares del *hardware* que pueden estar o bien en el estado 0 o bien en el estado 1 son las únicas «partes móviles» de la máquina, y qué máquina sea una computadora en cualquier momento dado depende de los ajustes de miles o de millones de pequeños elementos.

Una máquina virtual es lo que obtienes cuando le impones un tipo particular de instrucciones (más literalmente, disposiciones) a una máquina real que tiene una gran *plasticidad* —partes que se relacionan unas con otras y que pueden estar en muchos estados diferentes—. Como una máquina virtual hace trabajo informativo, puede *hacer el mismo trabajo* que una computadora cuyas «partes móviles» sean cambios de estado en el *hardware* si convierte todos esos cambios de estado en *representaciones* de las partes móviles. Puedes hacer una división larga con un lápiz en un pedazo de papel, o si eres realmente bueno con las divisiones puedes hacerla «en tu cabeza» tan sólo representando —imaginando— los trazos en una página o un pizarrón imaginarios. El resultado funciona igualmente bien de las dos maneras, pues es información: una

respuesta. Por el contrario, imaginar que preparas un sándwich de jamón si tienes hambre no puede remplazar la preparación de un sándwich de jamón. Las computadoras hacen tan bien el trabajo informativo «en sus cabezas» (representando la máquina que hace el trabajo) que puede ser prácticamente imposible saber si la máquina que estás usando y con la cual estás interactuando es una máquina procesadora de información para usos especiales «dedicada» y «con programas integrados» o una máquina virtual funcionando con un chip para todo uso. Por ejemplo, la mayoría, si no es que todos, los chips de computadora diminutos y baratos que ahora están integrados en elevadores, aparatos de aire acondicionado, tu coche, tu refrigerador y el control remoto de tu televisión son de hecho computadoras para todo uso que pueden ejecutar versiones de todos los programas que usas en tu computadora portátil, pero condenados a pasar toda su vida ejecutando un solo programa relativamente simple (el programa de control de arranque, el ciclo de descongelado, etc.) «quemado» en una memoria ROM que reduce toda su magnífica competencia a un solo truco, o dos. Esto es más barato que hacer chips para usos especiales diseñados para no hacer nada más que estas sencillas tareas.

El concepto de una máquina virtual es uno de los más útiles extensores de imaginación que hayan surgido de la informática, y después de haber demostrado su valor en ese dominio, está listo para exportarse a otras áreas. Uso el término en un sentido más o menos extendido (y en su momento explicaré por qué), así que vale la pena saber cuál es el significado original (y algunos dirían apropiado) del término. Los informáticos Gerald Popek y Robert Goldberg introdujeron el término.[7] En su uso original, significaba «un duplicado eficiente y aislado de una máquina real»: un duplicado hecho de… instrucciones. Una máquina real, llamémosla A, es *hardware* real, hecho de chips de silicio, cables, etc., y una máquina virtual es un programa de computadora (que se ejecuta en alguna máquina real *diferente,* B) que imita perfectamente el *hardware* de A: puede funcionar un poco más lento, porque tiene que componer todas las más básicas operaciones de A a partir de las operaciones básicas disponibles en su propio *hardware,* B, pero ejecuta todos los mismos programas. Si tienes un programa que se haya escrito para ejecutarse en el *hardware* A, debería ejecutarse sin problemas también en el *hardware* B cuando el *hardware* B esté ejecutando la imitación de máquina virtual de A.

Éste es un truco de lo más útil, y no sólo por el ahorro que propone: digamos que no tienes una Mac, pero tienes por ahí todo ese *software* valioso que sólo puede ejecutarse en Macs. Bueno, pues escribes una imitación de máquina virtual (MV) de una Mac que se ejecute en tu PC, y entonces tu PC podrá ejecutar todo el *software* para Mac cada vez que ejecute la MV de Mac. Tu PC «aparentará» ser una Mac, ¡y el *software* ni se dará por enterado! Piensa en alguien que se ha fracturado el brazo y lo tiene enyesado. El yeso restringe seriamente el movimiento de su brazo, y también el peso y la forma exigen ajustes en el resto de sus movimientos corporales. Ahora piensa en un mimo (digamos

Marcel Marceau) imitando a alguien con el brazo enyesado; si el mimo lo hace bien, sus movimientos corporales estarán restringidos exactamente de la misma manera: tiene un *yeso virtual* en el brazo, y será «casi visible». Una PC que imita una Mac ejecutando una MV de Mac debería ser indistinguible —para el *software* que se ejecuta en ella y para el observador externo— de una Mac real.

La realidad invierte mi ejemplo, más o menos. Aunque la gente ha desarrollado MV de Mac que funcionan en las PC, que yo sepa éstas son tretas más que *software* serio que pueda utilizarse. Las Mac, por otra parte, sí tienen una MV confiable y fácil de usar que ejecuta un sistema operativo para PC, Windows, lo que permite a los dueños de Mac usar el *software* para Windows PC que quieran. En la actualidad, la mayoría de los programas no se escriben para ejecutarse en un *hardware* particular: se escriben para ejecutarse en *sistemas operativos* particulares (que a su vez funcionan en diversos *hardwares*). Una de las razones para ampliar el concepto de máquina virtual es incluir imitaciones virtuales de sistemas operativos. Un sistema operativo es por sí mismo una clase de máquina virtual que permite que *hardwares* ligeramente distintos ejecuten los mismos programas, pero un sistema operativo es *sólo software:* no imita ningún *hardware* real sino que crea, de hecho por estipulación, una máquina imaginaria que sigue determinadas reglas, acepta determinados *inputs,* etcétera.

Otra razón para extender el concepto es el hecho de que una de las máquinas virtuales más populares y omnipresentes es la Máquina Virtual Java, o MVJ, que, como un sistema operativo, no es imitación de ninguna máquina de *hardware* real, sino que existe *sólo* como una máquina de *software.* Java es la invención que más ha hecho para que internet sea tan versátil como es; te permite descargar desde sitios web pequeños programas —*applets* de Java— con los que puedes resolver crucigramas, jugar Sudoku, explorar mapas, agrandar fotos, participar en juegos de aventura con otros jugadores en el otro extremo del mundo, y también usar la computadora para muchos asuntos más «serios». Un diseñador de sitios web puede escribir programas en el lenguaje de programación de Java sin tener que saber si las computadoras de los usuarios que vayan a ese sitio serán Mac o PC (o máquinas Linux), porque un *applet* de Java siempre se ejecuta en una MVJ específicamente diseñada para ejecutarse ya sea en una Mac o en una PC o en una máquina Linux. En pocos segundos se descarga e instala en tu computadora la MVJ indicada, y entonces el *applet* de Java se ejecuta en la MVJ, como magia. (Puede ser que hayas notado cuando las actualizaciones de Java se descargan a tu computadora, ¡o tal vez no lo has notado! Idealmente, puedes olvidarte de qué MVJ se instala en tu computadora y tan sólo esperar que cualquier sitio web al que vayas tendrá *applets* de Java que ya funcionan en tu MVJ, o si no, en ese momento instalará las actualizaciones de Java necesarias para que funcionen.)

En mi sentido amplio del término, pues, casi cualquier programa de computadora

puede considerarse una máquina virtual, pues es *software* —una lista sistemática de instrucciones— que, cuando se ejecuta, convierte una computadora de uso general en una máquina para usos especiales *que podría haber sido diseñada y armada como hardware.* Una de las más brillantes contribuciones de Alan Turing a la ciencia —de hecho a la civilización, desde mediados del siglo XX— es la idea de una computadora «universal» (lo que hoy en día llamamos una máquina de Turing universal), que puede *convertirse en* cualquier otra computadora claramente diseñada ¡con tan sólo instalar y ejecutar en ella un programa! (Por si decidiste saltártelo, esto se explica con detalle en el apartado 24.) No tienes que construir todas las diferentes computadoras de *hardware* imaginables: una sola es todo lo que vas a necesitar; el *software* hará todo el resto. Desde tiempos de Turing, hemos tenido la extraordinaria idea de que puedes agarrar un complicado pedazo de materia: el *hardware,* con mucha plasticidad: cajas o registros de «memoria» ajustables, y poner en ellas un conjunto de instrucciones que, cuando se ejecutan, convertirán tu pedazo de materia en cualquier computadora que puedas imaginar claramente.

Una máquina de Turing —o una computadora portátil— ejecuta una instrucción cada vez y pasa a la siguiente instrucción, pero podemos extender la idea a computadoras «paralelas» que pueden ejecutar muchas instrucciones (o muchos millones de ellas) cada vez. Un *registro* es cualquier lugar en el *hardware* que puede mantenerse en un estado u otro (como el estado cero y el estado uno de los bits en tu computadora, pero no necesariamente limitado a dos estados diferentes) hasta que algo le dice que cambie de estado. Cualquier sistema de registros que puedan realizar algunas operaciones elementales con base en esos estados (como cambiar un registro de un estado a otro, o usar el estado de un registro para determinar qué operación realizar después) puede tener sus registros ajustados para «calcular una función» o «ejecutar un programa». Así, cualquier *hardware* como ése puede ejecutar una máquina virtual diseñada para explotar estos pasos básicos. Y un truco que puede hacerse una vez puede hacerse dos o tres veces más, implantando máquinas virtuales en máquinas virtuales en máquinas virtuales en… *hardware.*

Piensa en un programa de ajedrez escrito en Common Lisp (un lenguaje de computadora de alto nivel) que funciona en Windows 7 (un sistema operativo) que se ejecuta en una PC. Esto es una PC que simula ser una máquina de Windows que simula ser una máquina Lisp que simula ser una máquina que juega al ajedrez. Si nos fijamos en los detalles del programa en el más alto nivel, serán más o menos comprensibles para un observador entendido en informática y en ajedrez («¡*Ajá!* Esta subrutina genera todas las respuestas legales al movimiento del alfil y luego activa la subrutina de evaluación que…») En contraste, mirar el código máquina real del mismo programa, las cadenas de ceros y unos con que se alimenta el registro de instrucciones del *hardware,* es una buena

manera de volverse loco o ciego, así que sabiamente dirigimos nuestra atención a los niveles más altos. En cada nivel *podemos* ver el bosque a pesar de todos los árboles porque los detalles del nivel inmediato inferior son convenientemente invisibles. El paralelo entre la cascada de máquinas virtuales en una computadora y la cascada de homúnculos de la visión de la mente que tiene el funcionalismo homuncular no es mera coincidencia. El espectacular éxito de las máquinas virtuales para ayudarnos a crear y comprender aplicaciones materiales de tareas hasta ahora inconcebibles (reservar boletos de avión, jugar al ajedrez, pronosticar el tiempo, tomar dictado, etc.) infunde la esperanza de poder lograr una proeza similar —sólo similar— cuando le hagamos ingeniería inversa al cerebro.

Entonces quizá las semejanzas entre los cerebros humanos cuyos dueños hablan francés, a pesar de todas las diferencias anatómicas observables, encuentran su mejor descripción en el *nivel de la máquina virtual:* todos los francófonos tienen una u otra versión de la MVF, la Máquina Virtual Francesa, un sistema hecho de disposiciones o microhábitos entrelazados que están de alguna manera almacenados en los miles de millones de registros del cerebro. Los cerebros de los hispanohablantes se distinguirán por un sistema de pautas confiables similar, la MVE, la Máquina Virtual Española. Si le dices a un francófono «*Donnez-moi le sel, s'il vous plaît*», puedes confiar en que la MVF regulará la misma conducta que obtienes de un hispanohablante cuando alimentas la MVE de su cerebro con el *input* «Pásame la sal, por favor». Ahora bien, ¿cómo construimos una MVF o una MVE que pueda ejecutarse en un cerebro?

Todavía no sabemos cómo describir esos diferentes niveles en las actividades de los cerebros de gente que juega al ajedrez o habla francés.[8] Sin duda, no habrá nada parecido a los mapeos precisos que han permitido a los programadores crear sus inventos en los más altos niveles con la total confianza de que del compilador (programa que recibe las instrucciones de alto nivel y las traduce a un código que puede ejecutarse en el *hardware*) surgirá un programa funcional. Sólo que ahora sí tenemos una prueba de concepto importante: conocemos al menos una manera de entender las competencias de alto nivel de una máquina con billones de partes móviles sin necesidad de invocar ningún **tejido mágico**.

26. Algoritmos

En *La peligrosa idea de Darwin* presenté esta manera de ver a lo que asciende la gran idea de Darwin:

> La vida en la Tierra se ha generado a lo largo de miles de millones de años en un solo árbol —el árbol de la vida— a partir de un proceso algorítmico u otro.[9]

¿Exactamente qué es un algoritmo entonces? De hecho, compiten entre sí varios sentidos del término, y el mío es tal vez el más amplio. Las páginas siguientes están tomadas de ese libro, con algunos ajustes.

Darwin descubrió el poder de un *algoritmo.* Un algoritmo es cierta clase de proceso formal en el que puede confiarse —lógicamente— para que arroje cierta clase de resultado siempre que es «ejecutado» o ejemplificado. Los algoritmos no son nuevos, ni lo eran en tiempos de Darwin. Muchos procedimientos aritméticos comunes, como hacer la división larga y sacar el saldo de tu cuenta de cheques, son algoritmos, y lo son también los procedimientos para jugar un «gato» perfecto y para poner en orden alfabético una lista de palabras. Lo que es relativamente nuevo —y nos permite una valiosa visión en retrospectiva del descubrimiento de Darwin— es la reflexión teórica de matemáticos y lógicos sobre la naturaleza y el poder de los algoritmos en general, un desarrollo del siglo XX que llevó al nacimiento de la computadora, lo que a la vez ha llevado a una comprensión mucho más profunda y vívida del poder de los algoritmos en general.

El término *algoritmo* llega al español medieval como *alguarismo,* de Al-Juarismi, sobrenombre del matemático persa cuyo libro sobre procedimientos aritméticos escrito en el siglo IX tradujo en el siglo xi al latín Adelardo de Bath o bien Robert de Chester. La idea de que un algoritmo es un procedimiento infalible y algo «mecánico» ha estado presente durante siglos, pero fue el trabajo pionero de Alan Turing, Kurt Gödel y Alonzo Church en la década de 1930 lo que más o menos fijó nuestra interpretación actual del término. Aquí destacaremos tres características clave de los algoritmos (las tres son un poco difíciles de definir).

1. *Neutralidad del sustrato.* El procedimiento para la división larga funciona igualmente bien con lápiz o pluma, papel o pergamino, escrita con luces de neón o con vapor arrojado desde un avión, y con el sistema de símbolos que desees. El poder del procedimiento se debe a su estructura *lógica,* no a los poderes causales de los materiales que se usan en su ejemplificación, siempre que esos poderes causales permitan que los pasos prescritos se sigan con exactitud.

2. *Simpleza subyacente.* Aunque el diseño general del procedimiento puede ser brillante, o arrojar resultados brillantes, cada uno de los pasos que lo componen, y la transición entre los pasos, es totalmente simple. ¿Qué tan simple? Lo suficiente para que un idiota diligente —o un sencillo dispositivo mecánico— lo lleve a cabo. La clásica analogía estándar indica que los algoritmos son una especie de *recetas* hechas para cocineros *novatos.* Un libro de recetas escrito para grandes *chefs* podría incluir la frase «Escalfa el pescado en un vino apropiado hasta que esté casi cocido», pero un algoritmo para ese mismo proceso podría empezar así: «Escoge un vino blanco que diga "seco" en la etiqueta; toma un sacacorchos y abre la botella; echa 3 centímetros de vino en el fondo de una cacerola; enciende el quemador debajo de la cacerola a un fuego alto…»: un tedioso desglose del proceso en pasos facilones, para los que no hacen falta sabias decisiones, cálculos delicados ni intuición alguna en el lector de la receta.
3. *Resultados garantizados.* Haga lo que haga un algoritmo, siempre lo hace si se ejecuta sin tropiezos. Un algoritmo es una receta infalible.

Es fácil ver cómo estas características hacen posible la computadora. *Todo programa de computadora es un algoritmo,* hecho de pasos simples que un sencillo mecanismo u otro pueden ejecutar con una tremenda fiabilidad. Los circuitos electrónicos son la elección habitual, pero el poder de las computadoras no les debe nada (salvo la velocidad) a las singularidades causales de los electrones precipitándose de un lado a otro en chips de silicio. Dispositivos con fotones que trajinan de aquí para allá en fibra de vidrio pueden realizar los mismísimos algoritmos incluso más rápido, o equipos de personas con lápiz y papel, mucho, mucho más lento.

Lo que Darwin descubrió no fue en realidad *un* algoritmo, sino más bien una gran clase de algoritmos relacionados entre sí que no tenía manera clara de distinguir.

27. Automatizar el elevador

Antes de dejar nuestro intermedio sobre computadoras quiero presentar otro grupo de ideas útiles, sobre el **código fuente**, el comentario y el código objeto, que tienen una aplicación importante para nuestro intento de entender cómo podría el significado residir en los cerebros. Es aconsejable estudiar en detalle un ejemplo muy simple a fin de afianzar los conceptos antes de abordar uno con el que se pueda uno quebrar la cabeza. (En el campo de la inteligencia artificial, éstos se conocen con el lindo nombre de *problemas de juguete*. Primero resuelves el problema de juguete antes de abordar el grande y difícil problema del mundo real.) Así, pues, ésta es una historia —inventada por simplicidad, pero realista por lo demás— sobre cómo los elevadoristas fueron remplazados por chips de computadora.

En mi juventud había elevadoristas, personas cuyo trabajo consistía en subir y bajar todo el día en un elevador, deteniéndose en los pisos indicados para que los pasajeros entraran y salieran. En los primeros tiempos manipulaban una curiosa manija que podía moverse en el sentido de las manecillas del reloj, o en el sentido contrario, para que el elevador subiera o bajara, y se necesitaba destreza para detenerlo justo a la altura correcta. A menudo la gente tenía que subir o bajar unos pocos centímetros al entrar o salir, y los elevadoristas siempre daban la advertencia. Tenían muchas reglas sobre qué decir y cuándo, y a qué pisos ir primero, y cómo abrir las puertas, etc. Su adiestramiento consistía en memorizar las reglas y luego practicar: seguir las reglas hasta hacerlo automáticamente. Las reglas mismas habían sido resultado de un trabajo asiduo a lo largo de los años en un proceso de diseño en el que se hicieron muchísimas revisiones y mejoras. Supongamos que este proceso prácticamente hubiera terminado y se tuviera como resultado un reglamento ideal. Funcionaba a las mil maravillas. Cualquiera que siguiera las reglas al pie de la letra era un magnífico elevadorista.

Ahora imagina qué pasó cuando se hizo posible que un simple programa de computadora asumiera todas las tareas de mando del elevadorista (de hecho, esto ocurrió gradualmente, cuando se fueron introduciendo diversos dispositivos mecánicos automáticos para quitarle al elevadorista las tareas más especializadas, pero vamos a imaginar que los elevadores pasaron de los elevadoristas a los sistemas completamente controlados por una computadora de sopetón).

El fabricante de elevadores, supongamos, convoca a un equipo de ingenieros de *software* —programadores— y les pasa el reglamento que los elevadoristas han estado siguiendo, diciéndoles: «*Esto* es una especificación del desempeño que queremos; hagan un programa de computadora que siga todas las reglas de este libro tan bien como los mejores elevadoristas, y con eso estaremos satisfechos». Mientras los programadores revisan el reglamento hacen una lista de todas las acciones que deben tomarse y las condiciones bajo las cuales son obligadas o están prohibidas. En el proceso pueden

arreglar un poco del desorden del reglamento. Por ejemplo, si incorporan sensores para garantizar que el elevador siempre se detenga exactamente donde debe, pueden eliminar el bucle que obliga al elevadorista a decir «Por favor, dé un paso hacia arriba» o «Por favor, dé un paso hacia abajo», pero pueden dejar intacto el «Piso [N], camine con cuidado» (en una grabación). Luego escriben un bosquejo del programa en lo que a menudo se llama seudocódigo, una clase de lenguaje híbrido que está a mitad de camino entre la lengua cotidiana y el sistema, más demandante, del código fuente. Una línea de seudocódigo podría verse más o menos así: «si llamapiso > pisoactual, entonces SUBE hasta llamapiso = pisoactual y ALTO; ABREPUERTA. ESPERA…»

Ya que el plan esté claro en seudocódigo, y parezca ser lo que se quiere, el seudocódigo puede traducirse a código fuente, que es un sistema de operaciones mucho más riguroso y estructurado e incluye definiciones de términos: variables, subrutinas, etc. Sigue siendo posible que los seres humanos descifren rápidamente el código fuente —después de todo, ellos lo escribieron—, y por lo tanto las reglas y términos del reglamento siguen ahí representados muy explícitamente, si sabes cómo buscarlos. Hay dos características que facilitan esto: en primer lugar, los nombres de las variables y operaciones suelen elegirse para demostrar el significado que buscan (llamapiso, sumapeso, dipiso, etc.). En segundo lugar, como vimos en el apartado 24, los programadores pueden agregar *comentarios* a su código fuente, explicaciones entre paréntesis que les dicen a otros lectores humanos del código fuente lo que el programador tenía en mente y qué es lo que se espera que hagan las diferentes partes. Cuando escribes un programa, es recomendable agregar comentarios para ti mismo sobre la marcha, pues no es difícil que se te olvide lo que pensabas que la línea de código hacía. Cuando regreses a corregir errores de programación, estos comentarios serán de lo más útiles.

El código fuente tiene que componerse con mucho cuidado y conforme a una sintaxis estricta, con cada elemento en el lugar indicado y toda la puntuación en el orden correcto, dado que con él tiene que alimentarse un programa *compilador* que toma el código fuente y *lo* traduce a las secuencias de las operaciones fundamentales (en *código objeto*) que la máquina real (o la máquina virtual) puede ejecutar. A un compilador no se le puede pedir que adivine qué quiso decir un programador con una línea de código fuente; el código fuente debe decirle al compilador exactamente qué operaciones realizar; pero el programa compilador puede tener muchas maneras distintas de realizar esas tareas y sabrá encontrar una eficiente para las circunstancias. Algunos compiladores son mejores que otros; si alimentas con el mismo programa (escrito en código fuente) dos diferentes compiladores, el código objeto que un compilador produzca puede ejecutarse significativamente más rápido que el otro código objeto, por ejemplo. Imagina que has escrito un programa de ajedrez y que alimentas dos diferentes compiladores con su

código fuente. Entonces haz que las dos versiones compiladas jueguen una contra otra en la misma computadora. Aunque las dos versiones «tienen los mismos pensamientos en el mismo orden» (debe ser así, pues tienen exactamente el mismo código fuente), una podrá ganarle siempre a la otra simplemente porque los piensa más rápido, usando menos ciclos máquina, y por lo tanto puede ver más lejos hacia delante en el tiempo disponible.

Regresemos a nuestro elevador. Una vez que el compilador ha compilado el código objeto, la máquina (virtual) puede *ejecutarlo* (es un archivo ejecutable, y su nombre normalmente tendrá el sufijo «.exe» —del inglés *execute*—). Puede ser que necesite rondas de depuración (habrá que regresar al código fuente, ajustarlo, volver a compilarlo, etc.), pero tarde o temprano será un producto «terminado». Se podrá «quemar» en un pequeño chip de memoria ROM que contiene una máquina universal —y cualquier cantidad de máquinas virtuales encima de ella— e instalar en el elevador. Instalarlo implica conectar las entradas de los transductores, como las señales de los botones, y de la báscula en el piso que pesa a los pasajeros y de otras partes, y conectar las salidas a los efectores (ejecutar los motores que abren y cierran la puerta, subir y bajar la cabina, y actualizar los letreros de la pantalla y reproducir las grabaciones). ¡*Tarán!* Una máquina ha remplazado a un ser humano real —no a un homúnculo figurativo—. Y la máquina *sigue las mismas reglas* que el elevadorista. ¿De verdad lo hace? ¿*De verdad?* Bueno, no. **Maomeno** sigue las mismas reglas. Esto es un buen ejemplo intermedio entre un ser humano que memoriza —y por lo tanto literalmente representa en su mente, y consulta — las reglas que dictan su conducta, y los planetas, cuyas órbitas son elegantemente descritas en ecuaciones que los planetas «obedecen». También los seres humanos a menudo ocupamos un nivel intermedio, cuando hemos internalizado o mediante la práctica hemos vuelto rutinario un conjunto de reglas explícitas que luego podemos descartar o incluso olvidar (como le pasa a un angloparlante con las reglas de pronunciación del inglés). Y también es posible **maomeno** seguir reglas que aún no se han hecho explícitas en forma depurada: las reglas de la gramática inglesa, por ejemplo, que siguen poniendo a prueba a los lingüistas. Para seguir con este ejemplo, hoy en día los lingüistas siguen jalándose los pelos tratando de escribir el reglamento para hablar un buen inglés, ¡mientras que todos los hablantes nativos del inglés a los 10 años ya tienen de algún modo instalada y depurada una versión bastante buena del código objeto de la MVI![10]

Antes de que dejemos este ejemplo, observa cómo los comentarios que se insertan en el código fuente para ayudar a los diseñadores del programa a llevar un registro de los propósitos de todas las partes entrelazadas del *software* no tienen una contraparte en el proceso de diseño que crea todo el *hardware, firmware* y *software* que caracteriza a nuestros cerebros. Cuando la selección natural instala diversas estructuras funcionales en nuestros cerebros, es como un código sin comentarios —están ahí por una razón, pero

éstas no se hallan representadas en la estructura con ningún letrero o explicación, que de todas formas el cerebro no podría entender (esto se discutirá con más detalle en el apartado 40)—. También hay razones no comentadas ni anunciadas para los ajustes que se realizan durante el desarrollo y el aprendizaje. Como los lingüistas, seguimos batallando para hacer ingeniería inversa de todas estas «reglas» y «procedimientos». La tarea es todavía más difícil que hacer ingeniería inversa del código objeto para recuperar código fuente (sin los comentarios), pero en principio puede hacerse.

Resumen

Durante cientos de años hemos tenido pruebas abundantes de que el cerebro es de alguna manera el asiento del alma, pero hasta mediados del siglo XX era completamente inimaginable cómo podía ser así. El cerebro podía verse como si estuviera compuesto de muchos órganos diferentes de formas curiosas, en pares izquierdos y derechos, y los primeros anatomistas les daban nombres gráficos: el hipocampo (caballo de mar), la amígdala (almendra), la arrugada corteza (como la de los árboles), pero ¿qué demonios hacían estas partes? No digerían la comida ni purificaban la sangre, ¿o sí? ¿Era el cerebro sólo un órgano para enfriar la sangre, una especie de radiador, como pensaba Aristóteles? Las partes estaban conectadas por fibras nerviosas, así que tal vez de alguna manera se comunicaban entre sí. Descartes sugirió que algunas fibras nerviosas eran como los hilos con que haces sonar las campanas: cuando jalabas un extremo, algo pasaba en el otro, pero ¿exactamente qué? Una campana sonando no te ayudaba en absoluto a comprender el cerebro como mente, y nadie tenía mejores ideas.[11]

Y luego llegó Turing a edificar sobre una tradición que se remontaba a Babbage, Pascal, Leibniz y otros, con la sugerencia de que un cerebro podría estar compuesto de partes sencillas que fueran a fin de cuentas sólo mecánicas (como una ratonera, una campana, una cerradura y una llave), pero si estas partes se organizaban para interactuar juntas de maneras más ingeniosas, en realidad podrían hacer algo inteligente por su cuenta, sin intervención humana ni fantasmas en la máquina que los dirigieran. Podrían hacer cómputos. Antes de que Turing diera con su idea, «computador» era el nombre de un oficio; la industria y el gobierno contrataban a miles de personas para que *fueran* computadores, tablas de cálculo para uso en los negocios, la navegación, la artillería y la banca, por ejemplo. A lo mejor, conjeturó, el cerebro es de suyo como un computador (humano); a lo mejor procesa información siguiendo obedientemente enormes listas de instrucciones muy simples (como *Inc* y *Dir*). Una vez que los primeros teóricos de la ciencia cognitiva, Alan Turing y John von Neumann; el creador de la cibernética, Norbert Wiener, y el creador de la teoría de la información, Claude Shannon, junto con muchos otros, articularon la idea, podría parecer hasta obviamente cierta: ¿cómo pudimos no haberlo visto todo este tiempo? Los cerebros deben asimilar información de los órganos de los sentidos y procesarla, computando de algún modo con ella, hasta haber extraído el valioso mineral del significado, que, con cómputos adicionales, podrían luego categorizar y almacenar para usarlo en el futuro en la dirección del comportamiento del cuerpo, que proporciona al cerebro su energía y su hogar protector. La innovación clave en la visión de Turing era que eliminaba ese elemento extraño que era demasiado evidente en las imaginaciones tempranas del procesamiento de información: coyunturas que requerían un dependiente, traductor o bibliotecario; que requerían, dicho brevemente, alguna clase de *entendedor* que apreciara la importancia de las señales. Turing vio que en un sentido esto

era ineludible: los procesos inteligentes siempre *requerirían* elegir una acción u otra con base en el discernimiento de alguna diferencia en la señal. No obstante, pudo reducir este entendimiento al mínimo indispensable: ramificación condicional, el proceso automático por el cual un dispositivo decide (**maomeno** decide) ir a la izquierda y no a la derecha porque percibe (**maomeno** percibe) 1 más que 0, A más que B, *x* más que *y*. Eso, y aritmética, era todo lo que necesitabas. Entonces podías fabricar dispositivos con cualquier nivel de discernimiento apilando máquinas virtuales arriba de máquinas virtuales arriba de máquinas virtuales —por decirlo anacrónicamente—. Esta concepción ha sido tentadora durante más de 50 años, pero los detalles, como ya hemos empezado a ver, no se están aclarando fácilmente. Si los cerebros son computadoras, no se parecen mucho a las computadoras que usamos todos los días. Necesitamos recordarnos las características fundamentales de las computadoras para entonces poder considerar alternativas más biológicamente realistas a las arquitecturas formales que aparecen en nuestros estereotipos.

El propósito de este intermedio ha sido aclarar esa concepción, y añadirle apenas el suficiente detalle, para que puedas *usarla* como herramienta de pensamiento, una prótesis de la imaginación que te ayude a entender los temas que ahora abordaremos: primero, un poco más sobre cómo podrían los significados residir en los cerebros (y en otras máquinas), y luego cómo la evolución podría diseñar esas ingeniosas arquitecturas sin ayuda de ningún Programador Maestro o Diseñador Inteligente. Entonces estarás listo para usar las herramientas de pensamiento adquiridas para pensar efectivamente sobre la conciencia y el libre albedrío, los temas más traicioneros que yo conozca.

[1] T. Wolfe, «Sorry, but Your Soul Just Died», en *Hooking Up,* Farrar, Straus & Giroux, Nueva York, 2000.

[2] H. Wang, «A Variant to Turing's Theory of Computing Machines», *Journal of the Association for Computing Machinery,* 4 (1): 63-92, 1957.

[3] Le agradezco a mi colega George Smith por haberme acercado a las máquinas de registro, en un curso de introducción a las computadoras que impartimos en Tufts a mediados de la década de 1980. Él reconocía el inmenso potencial pedagógico de las máquinas de registro y armó la estructura de exposición que empleo aquí, adaptada para un público un poco distinto. El Estudio de Software Curricular que George y yo fundamos en Tufts surgió de aquel curso.

[4] En el apéndice pueden encontrarse las soluciones a los ejercicios.

[5] ¡Esto significa que los demás ejercicios son obligatorios! Es en serio: si quieres que esta herramienta de pensamiento sea de provecho, tienes que practicar, practicar y practicar hasta que la domines. Resolver los sencillos ejercicios obligatorios podrá tomarte una o dos horas extras, pero vale la pena.

[6] Ada Lovelace, hija del poeta Lord Byron, fue una extraordinaria matemática y mucho más. En 1843 publicó su traducción de un comentario italiano sobre el Motor Analítico de Babagge, acompañada de sus propias notas, que eran más largas y profundas que el artículo que había traducido. Véase L. F. Menabrea, «Notions sur la machine analytique de M. Charles Babbage», *Bibliothèque Universelle de Genève*, 82 (41): 352-376, 1842. Entre estas notas estaba su cuidadosamente ideado sistema para emplear el Motor de Babbage para computar números de Bernoulli. Por esta razón es a menudo considerada la primera programadora de computadoras.

[7] G. J. Popek y R. P. Goldberg, "Formal Requirements for Virtualizable Third Generation Architectures", *Communications of the acm,* 17 (7): 412-421, 1974.

[8] La neurociencia cognitiva y la neurociencia informática se empeñan en entender esos niveles. La diferencia entre los dos subcampos es principalmente de énfasis: a los neurocientíficos informáticos les interesa más crear modelos (computarizados) de sus ideas que realmente funcionen, mientras que los neurocientíficos cognitivos muchas veces prefieren esbozar pautas de competencia e interacción de alto nivel que deben aplicarse en cualquier modelo de nivel más bajo. Durante años ha existido en las ciencias de la computación un estira y afloja similar: por un lado están los visionarios de la inteligencia artificial que no se molestan en escribir programas reales, sino que se conforman con demostrar hechos acerca de las especificaciones para cualquier programa exitoso que sirva para una u otra tarea; por el otro lado, ingenieros de hueso colorado que no se impresionan hasta no ver un código que de verdad funcione y cumpla con la tarea. Lo demás no es *software,* dicen con desdén: es *vaporware*. La impaciencia que muestran con sus colegas más especulativos no es nada comparada con la hostilidad que puede observarse entre diferentes niveles de neurocientíficos. Como me dijo una vez el director de un laboratorio (cuya especialidad eran los canales de calcio de los axones de las neuronas): «En nuestro laboratorio tenemos un dicho: si trabajas sobre una neurona, es neurociencia; si trabajas sobre dos neuronas, es psicología». ¡Y no lo decía como cumplido! Como la neurociencia cognitiva recibe la mayor parte de la atención mediática (a todo el mundo le fascinan los nuevos descubrimientos sobre ilusiones visuales, la memoria, la conciencia, el habla, el entendimiento, pero no todo el mundo se emociona con los cientos de distintos neuromoduladores y sus receptores, o con las interacciones astrocito-neurona… o con los canales de calcio), a menudo hay un alto grado de celos profesionales contaminando las actitudes de los neurocientíficos informáticos hacia sus vecinos de la neurociencia cognitiva.

[9] D. C. Dennett, *Darwin's Dangerous Idea…, op. cit.*

[10] Mi colega Ray Jackendoff, en *Patterns in the Mind: Language and Human Nature* (Harvester Wheatsheaf, Londres, 1993), lo llama la paradoja del aprendizaje de la lengua: de algún modo, los niños sin mayor esfuerzo absorben, y siguen, las reglas de la gramática que los lingüistas profesionales siguen buscando cómo expresar.

[11] De hecho, la gran imaginación de Descartes al menos le ayudaba a tener muy buenas ideas. Imaginó que los hilos al tirarse podrían abrir pequeñas puertas o poros que liberaran torrentes de «espíritus animales» acumulados (líquido cefalorraquídeo) que luego podrían hacer alguna clase de trabajo hidráulico; ¡esto no es mal

boceto para un amplificador! Mejor aún: vio que al menos algún trabajo inteligente (apto, apropiado) podía hacerse de manera enteramente mecánica con un dispositivo así: un *reflejo* automático. Retirar el pie del fuego cuando el calor jalaba un hilo era su ejemplo: pura maquinaria cerebral, ¡sin necesidad de mente!

V. Más herramientas sobre el significado

28. Algo con los pelirrojos

Ya hemos visto algunos problemas que acarrea la por otra parte tentadora idea de que toda la información que cargamos en nuestros cerebros —nuestras creencias, percepciones, recuerdos, principios, etc.— se puede descomponer en piezas parecidas a oraciones que se archivan y se quedan allí listas para ser usadas. La escritura cerebral no puede simplemente *instalar* una creencia falsa, y la gente puede compartir una creencia (acerca de aquel asesinato en Londres, por ejemplo) sin tener que compartir —aparentemente— una fórmula en *cerebrés*. Sin embargo, ¿qué *más* podría almacenar información en el cerebro? Los seres humanos podemos aprender cosas «por partes», así que debe de haber alguna manera de agregar datos independientes uno por uno, *a grandes rasgos*.

A los economistas (y otros) frecuentemente les gusta señalar que no puedes hacer una sola cosa. Hacer «una cosa» siempre tiene consecuencias. Del mismo modo, la idea de que puedes *aprender* sólo una cosa es discutible. Sin embargo, en una primera aproximación sí puedes. En estas páginas ya aprendiste que hay un mamífero que se llama pudú. A menos que hayas hecho una pausa para investigarlo, probablemente no puedas afirmar nada más acerca de los pudúes, aparte del hecho de que amamantan a sus crías, tienen columna vertebral y son relativamente raros (porque de otro modo seguramente habrías oído hablar de ellos antes). No es ningún misterio cómo aprendiste esto: leíste una oración y la creíste. Pero ¿pueden los animales o los niños pequeños que aún no tienen un lenguaje aprender *un solo hecho* (como un hecho expresado en una oración sencilla) a partir de una experiencia interesante? La idea de que el conocimiento, la creencia o el aprendizaje *tiene* que poderse descomponer en cachos del tamaño de una oración es quizá una ilusión antropomórfica. Los seres humanos nos topamos con muchos enunciados declarativos, hablados y escritos, a lo largo de un día, y de esa manera llegamos a conocer toda clase de hechos (y a adquirir nociones rudimentarias de falsedades). Algunos los almacenamos en bibliotecas y archivos, algunos nada más en nuestros cerebros. Rara vez memorizamos las oraciones palabra por palabra, pero cuando hacemos un guardadito de lo esencial de una oración con la que nos topamos, debe de ser —¿no?— asunto de almacenar una cosa u otra en algo así como una oración, una fórmula en *cerebrés*. Si no es así, ¿cuáles son las alternativas?

Supongamos que Pat dice que Mike «tiene algo con los pelirrojos». A grandes rasgos,

lo que Pat quiere decir es que Mike tiene un estereotipo de pelirrojo que es algo despectivo e influye en lo que Mike espera de los pelirrojos y de sus interacciones con ellos. No es sólo que Mike tenga un prejuicio contra los pelirrojos, sino que tiene algo un poco idiosincrásico y *especial* con los pelirrojos. Y puede ser que Pat esté en lo correcto —¡más de lo que él mismo supondría!—. Podría resultar que Mike sí tiene *algo,* no un *pensamiento,* una *creencia* ni una *imagen* ni ninguna de las otras cosas tradicionales que aporta nuestra experiencia consciente, sino un pedazo de maquinaria cognitiva **subpersonal** en su cerebro, que es *con los pelirrojos* en el sentido de que sistemáticamente entra en juego cada vez que se habla de los pelirrojos o de un pelirrojo, y ajusta varios parámetros de la maquinaria cognitiva de Mike que vuelve menos probable que albergue o confirme alguna hipótesis halagadora acerca de los pelirrojos, y más probable que tenga alguna conducta relativamente agresiva contra un pelirrojo que en otra circunstancia, etc. Ese *algo con los pelirrojos* podría ser muy complejo en su operación, o de lo más simple. La contribución del algo de Mike con los pelirrojos podría ser perfectamente determinada y también innegablemente significativa, y sin embargo ninguna *expresión* de este significado como una *oración creída verdadera* podría ser algo más que un nombre mnemotécnico para su función. Es decir, podría ser imposible caracterizar la función de este algo como una *creencia,* extrañamente específica o extrañamente vaga, de que *todos los pelirrojos son F*... (donde canjeamos «*F*» por el contenido que más justicia parezca hacerle a la actitud de Mike). Mike sin duda tiene una *actitud* hacia los pelirrojos, pero no es ninguna actitud *proposicional* específica, por usar la jerga filosófica. En otras palabras, se resiste a ser categorizada en el formato

Mike cree que: para todo *x, si x* es un pelirrojo entonces…,

sin importar cuán astutamente apilemos las cláusulas de exclusión, los cualificadores, las operaciones de probabilidad y otros ajustadores de contenido explícitos. Los filósofos (y otros teóricos) a menudo han tratado de «reducir» todos los estados cognitivos a estados que contienen información —llámense creencias y deseos— que puedan expresarse en fórmulas así, y si bien la táctica es una gran manera de ofrecer un bosquejo aproximado de la psicología de algunas personas (es la postura intencional, en efecto), la idea de hacerlo ultrapreciso es imposible. Podemos decir, si queremos, que hay diversas creencias *implícitas* en el sistema. Esto significa que el sistema está (actualmente) diseñado para operar «bajo la suposición» de que los pelirrojos del mundo tienen tales y cuales características. Cuando los programadores de computadora ponen un comentario en el código fuente para decirle a todo el mundo que este sistema se basa en un conjunto definido de suposiciones, saben que no tiene sentido dedicarle mucho tiempo a tratar de que las proposiciones sean precisas, pues se dan cuenta de que se trata de etiquetas mnemotécnicas para nosotros los observadores, nada que la computadora tenga que

maomeno leer y maomeno entender, e incluso para nosotros los observadores los comentarios no son *especificaciones* de contenido que puedan usarse como un químico usa fórmulas para describir moléculas. Dar una interpretación de cierta estructura subpersonal del cerebro desde el punto de vista de la postura intencional es como poner un comentario en unas cuantas líneas de código; cuando se hace bien, ofrece una etiqueta iluminadora, no una *traducción* al español o a ningún otro lenguaje natural ni una fórmula en cerebrés que el cerebro esté usando en su procesamiento de la información. Por perder de vista este truco, algunos filósofos han creado mundos fantásticos de maquinarias que manipulan oraciones internas, donde se imagina que hay una gran diferencia si el contenido de determinado suceso cerebral se expresa usando un predicado disyuntivo *(Veo un hombre-O-mujer)* o bien usando un predicado al que le falta estructura lógica *(Veo una persona),* por ejemplo.

¿Cuál es el propósito de esta bomba de intuición? Es simplemente un intento de sugerir que el conocido estribillo de apoyo al lenguaje del pensamiento —«¿Qué otra cosa podría ser?»— quizá tenga una buena respuesta que quitaría el viento que hinchaba las velas de aquellos a quienes les parece obvio. Me gustaría poder ofrecer una ambiciosa arquitectura informática alternativa que de manera triunfal mostrara una alternativa que funcionara, pero no puedo. Todavía nadie puede, pero por otra parte casi nadie está intentándolo, pues sigue difundiéndose ampliamente la convicción de que el lenguaje del pensamiento es «el único popote flotando», como alguien lo expresó hace años. Ten en mente, sin embargo, que nadie en la ciencia cognitiva ha desarrollado tampoco un prototipo del lenguaje del pensamiento, o siquiera ha hecho un gran esfuerzo para conseguirlo. Es un problema muy, muy difícil.[1] Quisiera alentar una mente abierta a este respecto.

29. El tragacuartos ambulante, la Tierra Gemela y el robot gigante

Puede parecer que al mencionar el algo de Mike con los pelirrojos y su código fuente en el mismo párrafo estoy alentando a los lectores a ignorar una grieta en los cimientos, de hecho una enorme brecha en medio de mi discusión de la intencionalidad: el problema de la *intencionalidad original*. John Searle acuñó el término, y la clara distinción que establece entre intencional *original* y *derivada* es, a primera vista, intuitivamente satisfactoria e incluso profundamente atractiva.[2] La doctrina de la intencionalidad original es la afirmación de que si bien algunos de nuestros artefactos pueden tener una clase de intencionalidad derivada de nosotros —nuestros libros y películas, nuestras computadoras e indicadores, por ejemplo—, nosotros tenemos una intencionalidad original (o intrínseca), absolutamente no derivada. Por ejemplo, las palabras impresas en esta página tratan de filosofía sólo porque nosotros, las personas que escribimos y leemos en español, tenemos pensamientos y creencias acerca de la filosofía que de alguna manera comunicamos usando estos rastros de tinta, los cuales no tratarían de nada sin quienes usamos las palabras. En contraste, nuestros pensamientos y creencias significan lo que significan independientemente de cualquier usuario ulterior: presentan una intencionalidad original, y en última instancia son la fuente de toda la intencionalidad derivada de muchos de nuestros artefactos. Se cuentan entre ellos no sólo palabras, oraciones y libros, sino mapas, películas, pinturas, signos, símbolos, diagramas y otras representaciones técnicas y, esto es importante, computadoras. Tanto una lista del mercado en un trozo de papel como una lista del mercado en tu iPhone son acerca de comestibles sólo en virtud del uso que haces de estas estructuras simbólicas, la interpretación que les otorgas, en beneficio de tu deseo de comprar comestibles y tu creencia de que el mercado es el lugar adonde ir, deseo y creencia que son acerca de comestibles más directa y originalmente. Aristóteles dijo que Dios es el Motor Inmóvil, y esta doctrina anuncia que somos los Significadores Insignificados.

Todos podemos estar de acuerdo con Searle en que nada tiene una intencionalidad *intrínseca* sólo en virtud de su forma física u otras propiedades por el estilo. Si, por una coincidencia alucinante, la forma

apareciera en rastros de distintos minerales en la pared de un acantilado en Marte, no sería («en sí mismo») un anuncio acerca de una bebida alcohólica, sin importar cuánta ilusión les hiciera a los lectores terrícolas interpretarlo así. La forma no trataría acerca de nada, a pesar de las primeras impresiones. Si algunos sucesos y objetos complicados del mundo tratan de otras cosas, debe de ser porque de alguna manera derivan la propiedad

de tratar de algo del hecho de estar al servicio de representar, interpretar sistemas intencionales cuyos estados (creencias, deseos, estados cerebrales) de cierto modo ya tienen intencionalidad.

¡Entonces la pregunta es si acaso *algo* tiene intencionalidad original! Y a primera vista podría parecer obvio que algo debe tener intencionalidad original, pues la intencionalidad derivada tiene que derivarse de algo. Y entonces los candidatos obvios para tener intencionalidad original serían las mentes humanas. Así que no sorprende que algunos filósofos muy eminentes, que por lo demás están en hondo desacuerdo con Searle en muchos temas —por ejemplo, Jerry Fodor y Saul Kripke—, concuerden con él, no obstante, en este aspecto. Ellos, y muchos del mismo parecer, piensan que las mentes humanas (o sus estados mentales) tienen una intencionalidad original y son a este respecto radicalmente diferentes de los sistemas de los robots.

Todos están rotundamente equivocados. Sí, equivocados. Lo digo en serio. Dado el innegable atractivo de la distinción entre intencionalidad original y derivada, cualquier intento de desacreditarla corre el riesgo de verse trastocado por una caridad que no viene al caso: «¡No *puede* estar diciendo en serio que estamos equivocados acerca de *eso*! ¡Debe de querer decir otra cosa, algo filosófico y esotérico que por falta de prudencia ha vestido con estos ropajes escandalosamente provocativos!» Tal vez la mejor manera de convencer a la gente de que realmente lo digo en serio sea sacar a relucir el ejemplo más vívido y claro de intencionalidad *derivada* que pueda encontrar, y entonces mostrar que el adorado contraste entre ese ejemplo y las mentes humanas como ejemplos de intencionalidad *original* se evapora cuando se examina más de cerca. Es mucho pedir, pero aquí va. Voy a tener que ligar tres bombas de intuición para conseguir la hazaña.

El tragacuartos ambulante

Piensa en una máquina estándar expendedora de refrescos, diseñada y construida en Estados Unidos, y equipada con un dispositivo transductor para aceptar y rechazar monedas de un cuarto de dólar. Llamémosla *tragacuartos.*[3] Normalmente, cuando se introduce una moneda de 25 centavos de dólar en un tragacuartos, éste entra en un estado, llamémoslo C, que «significa» (obsérvense las comillas de relieve para indicar que más bien *maomeno* significa) «Yo ahora percibo/acepto un auténtico cuarto de dólar». Estos tragacuartos son muy listos y sofisticados, pero difícilmente infalibles. Sí «cometen errores». Es decir, de manera no metafórica, a veces entran en el estado C cuando se inserta en ellos una ficha o algún otro objeto extraño, y a veces rechazan cuartos de dólar perfectamente legales: no entran en el estado C cuando *se esperaría que* entraran en él. Seguramente se pueden detectar ciertas regularidades en estos casos de «percepción equivocada», y seguramente alguien con el suficiente conocimiento de las leyes físicas pertinentes y de los parámetros de diseño de la maquinaria transductora de los tragacuartos podría predecir al menos algunos de los casos de «identificación equivocada». Podría deducirse muy directamente de diversas aplicaciones de leyes físicas que no sólo los cuartos de dólar sino también objetos de alguna clase K detonan el estado C, pero objetos de la clase J (demasiado pesados) o de la clase L (magnéticos, a diferencia de los cuartos de dólar) no. Entonces los objetos de la clase K serían buenas *fichas,* que confiablemente «engañan» al transductor. (Mira cuántas veces he usado el operador maomeno en este párrafo para poder usar la postura intencional al darte las especificaciones del tragacuartos. Trata de reescribir el párrafo sin usar la postura intencional y te darás cuenta de cuán eficiente ésta es, y de cómo el operador maomeno puede ser prácticamente indispensable para tales propósitos.)

Si los objetos de la clase K se hicieran más comunes en el ambiente normal del tragacuartos, esperaríamos que los dueños y diseñadores de estas máquinas desarrollaran transductores más avanzados y sensibles que confiablemente pudieran discernir entre genuinas monedas de un cuarto de dólar y fichas de la clase K. Desde luego, podrían hacer entonces su aparición falsificaciones más elaboradas, y se requerirían mayores avances en los transductores detectores, y en algún punto esta escalada de ingeniería alcanzaría rendimientos decrecientes, pues no hay tal cosa como un mecanismo *infalible.* Mientras tanto, sería prudente que ingenieros y consumidores se conformasen con los tragacuartos estándar y rudimentarios, pues no es rentable protegerse de abusos insignificantes.

Lo único que hace que el artefacto sea un detector de cuartos de dólar y no de fichas, ni un detector de cuartos *o* fichas, es la intención compartida por sus diseñadores, fabricantes, dueños y consumidores. Sólo en el ambiente o contexto de esos usuarios y sus intenciones podemos señalar algunas de las ocasiones del estado C como «verídicas»

y otras como «equivocadas». Sólo en relación con ese contexto de intenciones podríamos justificar llamar a ese artefacto un tragacuartos, para empezar.

Me imagino que hasta aquí tengo a Searle, Fodor, Kripke y otros moviendo la cabeza en señal de asentimiento: exactamente así es con esos artefactos; éste es *un ejemplo típico de intencionalidad derivada* al descubierto. Entonces a nadie le avergüenza admitir que determinado tragacuartos, directamente salido de la fábrica estadunidense y con una etiqueta que dice «Tragacuartos modelo A», podría instalarse en una máquina de refrescos panameña, donde empezara a desquitar su sueldo como aceptador y rechazador de monedas de un cuarto de balboa, de curso legal en Panamá, y fácilmente distinguible (por seres humanos) de los cuartos de dólar estadunidense por los dibujos y la escritura impresos en ellas, pero no por su peso, grosor, diámetro o composición material.

No estoy inventándolo. Sé de buenísima fuente —Albert Erler de la tienda numismática Flying Eagle Shoppe, Rare Coins— que las máquinas expendedoras estándar no pueden distinguir entre los cuartos de dólar estadunidense y los cuartos de balboa panameño acuñados entre 1966 y 1984. No es de extrañar, pues estos últimos se acuñaron con reservas de cuartos de dólar en casas de moneda estadunidenses. Y —para satisfacer a los curiosos, aunque para el ejemplo es estrictamente irrelevante— al tipo de cambio de hoy (2011) un balboa equivale a 0.98 dólares, así que un cuarto de balboa es, hoy, ligerísimamente menos valioso que un cuarto de dólar.

Normalmente ese tragacuartos, mandado a Panamá, seguiría entrando en cierto estado físico —el estado con las características físicas con que solíamos identificar el estado C— cada vez que un cuarto de dólar estadunidense, un objeto de la clase K o un cuarto de balboa panameño se insertara en él, pero ahora un conjunto diferente de tales ocasiones cuentan como errores. En el nuevo ambiente, los cuartos de dólar cuentan como fichas, como objetos de la clase K, como inductores a error, a percepción o representación equivocadas. Después de todo, de vuelta a Estados Unidos un cuarto de balboa es una especie de ficha.

Una vez que nuestro tragacuartos reside en Panamá, ¿debemos decir que sigue presentándose el estado C? El estado físico en el que el dispositivo «acepta» monedas sigue presentándose, pero ¿ahora no deberíamos más bien decir que identificamos ese estado físico como la «realización» de un nuevo estado, CB? Bueno, hay una considerable libertad sobre lo que deberíamos decir, pues después de todo un tragacuartos es sólo un dispositivo, y hablar acerca de sus percepciones y de sus percepciones equivocadas, sus estados verídicos y no verídicos —en suma, de su intencionalidad— es «sólo una metáfora». El estado interno del tragacuartos, como quieras llamarlo, *realmente* (originalmente) no significa ni «cuarto de dólar estadunidense aquí y ahora» ni «cuarto de balboa panameño aquí y ahora». *Realmente* no significa

nada —en eso harían hincapié Searle, Fodor y Kripke *(inter alia)*—. Su estado interno sólo maomeno significa algo, pero con eso bastará para plantear algunos problemas que también pueden presentársenos a nosotros, los que gozamos de intencionalidad original. Veamos los detalles.

El tragacuartos se diseñó originalmente para ser un detector de cuartos de dólar estadunidense. Ésa era su «función propia» y, muy literalmente, su razón de ser.[4] Nadie se habría molestado en inventarlo si no se le hubiera ocurrido este propósito. Y dado que este hecho histórico acerca de su origen autoriza cierta manera de hablar, ese dispositivo puede ante todo o debidamente caracterizarse como un tragacuartos, una cosa cuya función es detectar cuartos de dólar, para que *con respecto a esa función* podamos identificar tanto sus estados verídicos (cuando no se equivoca) como sus errores.

Esto no impediría que se arrancara a un tragacuartos de su nido y se le obligara a dar servicio con un nuevo propósito —cualquier nuevo propósito que las leyes de la física certificaran que pudiera cumplir confiablemente—: como un detector de K, un detector de cuartos de balboa, un tope para puertas, un arma mortal. En su nuevo papel podría haber un breve periodo de confusión o indeterminación. ¿Cuántos antecedentes tiene que acumular algo para ya no ser un tragacuartos sino más bien un detector de cuartos de balboa (un detector-cb), un tope para puertas o un arma mortal?[5] En su debut como detector-cb, tras diez años de fiel servicio como tragacuartos, ¿su estado C ya es una detección *verídica* de un cuarto de balboa, o podría haber una especie de error nostálgico por la fuerza de la costumbre, una identificación equivocada de un cuarto de balboa *como* un cuarto de dólar?

Tal como se ha descrito, el tragacuartos difiere notablemente de nosotros en que no está provisto de memoria de sus experiencias pasadas —ni siquiera *maomeno* memoria de sus *maomeno* experiencias pasadas—. Con todo, esto podría proveérsele fácilmente si se pensara que serviría de algo. Para empezar con la manera más fácil de adentrarse en este tema, supongamos que el tragacuartos (para referirnos a él con el nombre con que fue bautizado) está equipado con un contador que tras 10 años de servicio marca 1 435 792. Supongamos que no se vuelve a poner en cero durante su vuelo a Panamá, de modo que en su debut allá el contador se pone en 1 435 793. ¿Esto hace que el apoyo se incline a favor de la afirmación de que todavía no ha cambiado a la tarea de identificar correctamente cuartos de balboa? (A fin de cuentas, **maomeno** clasifica mal el suceso como uno de esos sucesos C —detecciones de cuartos de dólar estadunidense— que estaba diseñado para detectar.) ¿Las variaciones y complicaciones sobre este tema llevarían tus intuiciones en diferente dirección? (Gira todas las perillas de la bomba de intuición y ve qué pasa con tus intuiciones.)

Podemos asegurarnos que nada *intrínseco* acerca del tragacuartos considerado restringidamente por sí solo e independientemente de su historia anterior pueda

distinguirlo de un genuino detector-cb, hecho especialmente por encargo del gobierno panameño. Aun así, dado su linaje, ¿no hay un problema con su función, su propósito, su significado, la primera vez que entra en el estado que estamos tentados de llamar C? ¿Es éste un caso de entrada en el estado C (con el significado de «cuarto de dólar estadunidense aquí y ahora») o de entrada en el estado CB (con el significado de «cuarto de balboa panameño aquí y ahora»). Yo diría (junto con Millikan)[6] que el que su debut panameño cuente como entrar en el estado C o en el estado CB depende enteramente de si, en su nuevo nicho, fue *elegido por* su capacidad de detectar cuartos de balboa —literalmente elegido por el titular de la franquicia panameña de Pepsi-Cola, por ejemplo —. Si fue así, entonces, aunque sus nuevos propietarios podrían haberse olvidado de poner el contador en cero, su primer acto «perceptual» contaría como una identificación correcta por un detector-cb, pues eso es *para* lo que *ahora* serviría. Habría adquirido la detección de cuartos de balboa como su función propia. Si, por el otro lado, se hubiera enviado el tragacuartos a Panamá por error, o si hubiera llegado por pura coincidencia, su debut no significaría nada, aunque acaso pronto —incluso inmediatamente— las autoridades competentes (las que podrían obligarlo a dar servicio en un nuevo papel) reconocerían y apreciarían su utilidad, y acto seguido sus posteriores estados contarían como expresiones de CB. No obstante, antes de que fuera elegido para el puesto, sin importar qué tan bueno fuera para detectar cuartos de balboa, su estado de aceptación no *significaría* (a su derivada y *maomeno* manera, propia del artefacto que es) «cuarto de balboa panameño aquí y ahora». Cabe suponer que Searle y sus colegas estarían conformes de dejarme decir esto, pues, después de todo, el tragacuartos no es más que un artefacto. No tiene intencionalidad original, así que no hay ningún hecho «más profundo» que pudiéramos tratar de dejar al descubierto. Dirían que se trata sólo de una cuestión pragmática de la mejor manera de hablar cuando se habla metafórica y antropomórficamente sobre los estados del dispositivo.

Ahora que dominamos la intencionalidad *derivada,* veamos qué es lo que se supone que tiene de diferente la intencionalidad no derivada, *original,* es decir, *nuestra* intencionalidad. Es aquí donde Searle, Fodor, Kripke y muchos otros discrepan no sólo conmigo, sino con los filósofos Ruth Millikan y Paul y Patricia Churchland, los cognitivistas Douglas Hofstadter y Marvin Minsky, y prácticamente el resto del mundo que trabaja en el campo de la inteligencia artificial. Después de reñir más de 30 años, los ánimos siguen calentándose. ¿Cuál es, entonces, el tema en cuestión?

Supongamos que un ser humano, Jones, al mirar por una ventana piensa que ve un caballo. Puede haber, o no, un caballo allá fuera para que él lo vea, pero el hecho de que se encuentre en el estado mental de pensar que ve un caballo no es *en absoluto* un asunto abierto a interpretación (dicen Searle y compañía); es un hecho bruto, un ejemplo de *intencionalidad original.* Veamos qué pasa, entonces, si elaboramos un experimento mental paralelo al del embrollo panameño (pista: éste será un argumento de reducción al absurdo). Supongamos entonces que el planeta Tierra Gemela fuera exactamente igual que la Tierra, excepto porque tiene «chaballos» donde nosotros tenemos caballos.[7] Los chaballos parecen caballos, y los únicos que pueden distinguirlos son biólogos adiestrados equipados con *kits* de prueba de ADN, pero en realidad los chaballos no son más caballos que los delfines peces. Los tierragemelícolas llaman a los chaballos «caballo», «cheval» o «Pferd» (recuerda, la Tierra Gemela es exactamente igual a la Tierra excepto porque tiene chaballos).

Supongamos que despachamos a Jones a la Tierra Gemela, tierra de chaballos, sin que se dé cuenta (lo drogamos para el viaje y disponemos todo para que se despierte en la contraparte de su cama en la Tierra Gemela). Cuando luego lo ponemos frente a un chaballo, naturalmente dirá, y pensará, «¡Mira! Un caballo». Cuando esto pasa, o bien sigue en el estado antes provocado de creer que ve un caballo (una creencia equivocada, no verídica) o bien ese chaballo lo provoca a creer, por primera vez en su vida (y verídicamente), que está viendo un chaballo. ¿Cuál de las dos es, y cómo podríamos saberlo? ¿Es verdadera o falsa su creencia provocada por el chaballo? Si su primer pensamiento fue una creencia equivocada de que acababa de ver un caballo, ¿cuánto tiempo le tomaría, viviendo entre chaballos y hablando con los tierragemelícolas sobre chaballos, ajustar el *significado* del sonido «caballo» en su idioma (¡sin darse cuenta!)? Si tuviera hijos en la Tierra Gemela, ¿*su* palabra «caballo», que aprendieron en el regazo de su padre, significaría caballo o chaballo? Recuerda que nunca han visto un caballo: sólo chaballos.

Obviamente esto es un ejemplo estrafalario y extremo, pero plantea un tema interesante: ¿qué establece lo que nuestros términos significan, y cómo? ¿La historia cuenta para todo, y siempre, o el uso actual puede superar o invalidar la historia? En este caso, Jones no tiene ninguna percepción privilegiada que ofrecernos; él no tiene idea de que ya no está en la Tierra, así que imagino que insistiría en que su palabra «caballo» significa *caballo.* Su palabra, la palabra que sale de sus labios, deriva su significado de su creencia perceptual, y él sabe lo que cree: está viendo un caballo. Por eso dijo «¡Mira! Un caballo» (y quizá añadiría: «Nada podría ser más obvio»). Imaginemos entonces que le contamos de su viaje y de la sutil pero importante diferencia entre los caballos y los chaballos. ¿Qué diría entonces, o qué debería decir, y sobre todo, hay una buena razón

para pensar que cualquier cosa que diga sería decisiva? ¿No está haciendo lo que el resto de nosotros haríamos, teorizar acerca de un caso en el que ni nosotros ni él tenemos información privilegiada? Supón que dice que su palabra «caballo» *ahora* significa chaballo: cuando ve un chaballo y lo llama caballo, no está cometiendo un error. Ya sabes: a la tierra que fueres…

¿No puede simplemente declarar lo que significan sus palabras y con eso dar por terminado el asunto? ¿Y qué tal si más adelante olvidara lo que había declarado? A veces hacemos cosas así: «¡A partir de hoy, cuando diga "yobyob" querré decir "sal de mesa"! ¡Por favor, pásame la yobyob!» En el contexto de las teorías científicas, este tipo de definiciones estipulativas son una práctica importante y establecida, pero dependen de una comunidad cooperativa de comunicadores. Si Jones tuviera intencionalidad original, podemos suponer que entonces debería haber un hecho acerca del significado de sus términos bajo cualquier circunstancia, pero parece que Jones mismo no puede consultar su propia intencionalidad original mejor que nosotros los observadores externos. Supongamos, por ejemplo, que le mentimos a Jones cuando le decimos que lo llevamos a la Tierra Gemela, y él cree nuestras mentiras (en las bombas de intuición de los filósofos la gente puede ser terriblemente crédula). Si entonces nos dice que su palabra «caballo» ahora significa chaballo, ¿estará en lo cierto? A lo mejor lo que debería decir es que no tiene idea de qué signifique ahora su palabra «caballo». Entonces, sin embargo, como, que nosotros sepamos, también pudimos haber sido despachados a la Tierra Gemela, ¿no deberíamos todos reconocer que tampoco nosotros tenemos idea de lo que queremos decir con «caballo»?

Los que dudamos de toda la idea de *intencionalidad original* tenemos listas las respuestas para todas estas preguntas, pero hará falta un tercer experimento mental que sea lo suficientemente claro para tener oportunidad de ganarle a la intuición tradicional (así que ¡*caveat lector!,* voy a tratar de engatusarte para que abandones una valiosa corazonada).

El robot gigante

Imagina que decides que quieres experimentar la vida en el siglo XXV, y supón que para mantener vivo tu cuerpo tanto tiempo la única manera conocida fuera colocarlo en una especie de dispositivo de hibernación, donde descansaría, desacelerado y en coma, todo el tiempo que se quisiera. Podrías hacer los arreglos necesarios para meterte en la cápsula, que te pusieran a dormir, y que luego automáticamente se te despertara y dejara salir en 2401.

Diseñar la cápsula misma no es tu único problema de ingeniería, pues la cápsula debe estar protegida y tener el suministro necesario de energía (para refrigeración o lo que sea) durante casi 400 años. No podrás contar con tus hijos y nietos para esta administración, pues para 2401 llevarán mucho tiempo muertos y no puedes suponer que tus descendientes más lejanos, si es que los tienes, tendrán un vivo interés en tu bienestar. Así, pues, debes diseñar un supersistema que proteja tu cápsula y le proporcione la energía que necesite durante 400 años.

Hay dos estrategias básicas que considerar. En una de ellas, deberías encontrar la ubicación ideal, la mejor que puedas prever, para una instalación fija que tendrá el suministro necesario de agua, luz solar y cualquier otra cosa que tu cápsula (y el supersistema mismo) necesite en todo ese tiempo. El principal inconveniente para una instalación, o «planta», como ésa, es que no puede moverse si algún daño se aproxima: si, por ejemplo, alguien decide construir una autopista justo donde se encuentra. La segunda alternativa es mucho más sofisticada, pero evita este inconveniente: diseñar una instalación móvil para tu cápsula, junto con los sensores necesarios y los dispositivos que puedan alertarla con antelación para que pueda alejarse del daño y tratar de conseguir nuevas fuentes de energía según le hagan falta. En dos palabras: construye un robot gigante e instálale la cápsula (contigo adentro).

Evidentemente estas dos estrategias básicas están copiadas de la naturaleza: corresponden, a grandes rasgos, a la división entre plantas y animales. Una tercera opción de la naturaleza, una espora o semilla curtida que pueda sobrevivir indefinidamente adentro de su armadura, no está disponible para ti, pues la máquina que mantiene tus constantes vitales tiene grandes demandas de energía, y las esporas son lo más inerte y poco necesitado de energía que la naturaleza puede producir. Como la estrategia animal conviene a nuestros propósitos, supondremos que decides construir un robot para albergar tu cápsula. Deberías tratar de diseñarlo de modo que por encima de todo «elija» acciones que propiciarán lo que más te conviene. Los malos movimientos y las vueltas equivocadas tenderían a incapacitar a tu robot para cumplir el papel de protegerte hasta 2401, que es su única razón de ser. Está claro que esto es un dificilísimo problema de ingeniería que exige el más alto nivel de pericia para diseñar un sistema de «visión» que dirija su locomoción, y otros sistemas «sensoriales». Como tú estarás en

coma todo ese tiempo y por ende no puedes dirigir y planear sus estrategias, tendrás que diseñarlo de modo que genere sus propios planes en respuesta a las circunstancias cambiantes. Debe «saber» cómo «buscar», «reconocer» y luego «explotar» fuentes de energía, cómo mudarse a un territorio donde puedas estar más a salvo, cómo «prever» y luego «evitar» peligros. Con tanto por hacer, que además debe hacerse rápido, lo mejor será que economices siempre que puedas: no le des a tu robot más habilidad de discernimiento que la que probablemente necesite para distinguir lo que tenga que distinguir en el mundo.

Una vez más observa que he puesto todos los términos intencionales o «mentalistas», como «sensoriales», «buscar» o «prever», entre comillas de relieve, para indicar que se trata de una clase específica de maomeno intencionalidad, intencionalidad *derivada,* intencionalidad que depende enteramente de tus propósitos humanos. Éste es *tu* artefacto, y sea cual sea su intencionalidad, te la debe a ti, su autor. Si no hubiera puesto las comillas, podría acusárseme de meter furtivamente alguna ideología, o de confiar en el hecho de que los ingenieros y otros tienen la costumbre de usar ese lenguaje —sin comillas— para hablar de las especificaciones de un dispositivo que maneja información (como el controlador del elevador). A propósito *no* estoy haciendo esa movida; por mor del argumento, estoy concediendo que *cualquier* uso de lenguaje intencional para describir o prescribir talentos en los artefactos es puramente metafórico. Observa también que, como el tragacuartos, la maquinaria del robot está sujeta a consideraciones económicas: necesita «detectar» o «discernir» muchas cosas, pero sus «discernidores» no serán infalibles. Puede cometer errores, pero *qué cuente como error depende finalmente de las necesidades y deseos de su autor.* Si el autor quisiera hacer un gracioso payaso robótico que fingiera andar metiendo la pata por el mundo «equivocando» las cosas, algunos de estos «errores» serían en realidad aciertos, triunfos de alguien que domina el oficio de payaso.

Regresemos a la bomba de intuición: dificultará tu tarea el hecho de que no puedes contar con que tu robot sea el único robot de los alrededores con esa misión. Si tu capricho se pone de moda, tu robot de pronto puede verse compitiendo con otros (y con tus descendientes humanos) por las reservas limitadas de energía, agua fresca, lubricantes y cosas por el estilo (véase el apartado 67 para una breve discusión de la importancia de la presencia de otros agentes). No cabe duda de que sería conveniente diseñar tu robot con un sistema de control lo bastante sofisticado para que le permitiera calcular los beneficios y riesgos de cooperar con otros robots, o de formar alianzas para mutuo beneficio, pero una vez más, un cálculo así tiene que ser una aproximación «rápida e imperfecta» arbitrariamente truncada por la presión del tiempo.

El resultado de este proyecto de diseño sería un robot que pudiera presentar alguna clase de autocontrol, pues es necesario que cedas a tu artefacto una parte de la dirección

más precisa y en tiempo real cuando ya te hayas ido a dormir. Como tal, el robot podrá derivar sus propios objetivos secundarios a partir de su evaluación del estado actual y de la importancia de ese estado para su objetivo último (que sigue siendo preservarte). Estos objetivos secundarios pueden alejarlo mucho si se embarca en proyectos que duren siglos, desacertados algunos de ellos, no obstante tus empeños durante la fase de diseño. Tu robot podría embarcarse en acciones contrarias a tus propósitos, incluso suicidas, si otro robot lo «convenciera», quizá, de subordinar su propia misión de vida a alguna otra.

Observa también que en este punto, aunque todos los estados y actos intencionales del robot se derivan de tus propósitos, empiezan a distanciarse un poco de ellos, digamos. Como diseñaste al robot para que «pensara por sí mismo» hasta cierto punto, su «pensamiento» puede escaparse de tus límites previstos. Para un ejemplo no ficticio y de la vida real de un artefacto así, piensa en una computadora jugadora de ajedrez que puede ganarle a su creador. Es cierto que la única razón por la que podemos decir que la computadora está actualmente «investigando» las opciones de enroque largo y «decidiendo» no enrocar es que es un artefacto diseñado por un artífice humano para hacer exactamente ese tipo de cosas. Sin embargo, también es cierto que, dado el objetivo del artífice consistente en hacer una computadora que juegue *bien* al ajedrez, muchas de las decisiones del artífice concernientes a *acerca* de qué (derivadamente) son los estados de la computadora le son impuestas al artífice: dado el hecho de que un jugador de ajedrez necesita información exacta sobre las reglas y el estado del juego, *debe* haber estados concernientes a cada alfil y a cada peón, y estados que impliquen una evaluación del juego si la dama de la computadora captura el caballo del oponente en la jugada actual, etc. Ningún decreto autoral podría hacer que un estado de la computadora tuviera que ver (derivadamente) con la cantidad de peones que quedan en el tablero si ese estado no estuviera adecuadamente ligado a la localización de todos y cada uno de los peones del tablero. Una vez que se ha definido el principal objetivo de un diseñador (haz un jugador de ajedrez, haz un robot gigante, haz un simulador de huracanes), la cruel naturaleza toma el mando y establece lo que va a funcionar y lo que no, y por lo tanto qué estados de qué sistema cuentan como equivocados o inexactos. Puede ser que los poetas se permitan declarar que un poema que parece tratar de caballos en realidad es sobre profesores —William Blake nos dice que «los tigres de la ira son más sabios que los caballos de la instrucción»—, pero los ingenieros en computación no pueden imponer así sus intenciones sobre sus creaciones.

Hagamos un balance. El simulacro de estados mentales del robot gigante sería exactamente eso: no será *realmente* decidir, ver, dudar y planear, sino *como si* decidiera, viera, dudara y planeara. Deberíamos hacer una pausa para asegurarnos de que entendemos lo que esta afirmación engloba. Desde luego que el robot que hemos imaginado es mucho más sofisticado que el humilde tragacuartos; le hemos otorgado el

poder de «planear» nuevos cursos de acción, «aprender» de los errores pasados, «hacerse de allegados» y «comunicarse» con sus competidores. Es más, para toda esta «planeación», «aprendizaje» y «comunicación», será necesario proveerlo de estructuras de control con un gran poder de autorreflexión o de autoseguimiento. En otras palabras, deberá tener un acceso como de humano a sus propios estados internos y tener la capacidad de «informar», «confesar» y «comentar» en torno a lo que él «considere» la importancia de sus propios estados internos (cuando «decida» que no «quiere» «mentirnos»). Tendrá «opiniones» sobre lo que significan estos estados, y sin lugar a dudas deberíamos tomarnos en serio esas «opiniones» como muy buenas pruebas —probablemente las mejores que podamos obtener— de lo que esos estados «significan» *metafóricamente hablando* (recuerda: sólo es un artefacto y no tiene intencionalidad original; estamos considerando su intencionalidad *derivada,* que para los observadores no será más obvia que la intencionalidad de nosotros los agentes «reales»). Al tragacuartos no se le concedió esta capacidad de influir sobre nuestros juicios interpretativos emitiendo, con aparente confianza, «declaraciones» de no tener ni idea de estar en Panamá o de haberse sorprendido al enterarse de la existencia de los cuartos de balboa.

Hay varias maneras como podría responderse a esta bomba de intuición, y pronto las estudiaremos, pero primero quiero señalar la implicación más sorprendente que se puede extraer de habernos mantenido firmes en nuestra suposición inicial: ningún artefacto, por mucha hechicería de inteligencia artificial con que se haya diseñado, tiene ninguna intencionalidad que no sea derivada. Si nos aferramos a esta postura, no podemos sino concluir que nuestra propia intencionalidad es exactamente como la del robot, pues la historia de ciencia ficción que he relatado no es nueva: es una variante de la visión de Richard Dawkins sobre nosotros y todas las demás especies biológicas como «máquinas de supervivencia» diseñadas para alargar el futuro de nuestros genes egoístas.[8] Somos artefactos, diseñados a lo largo de millones de años como máquinas de supervivencia para genes que no pueden actuar veloz e informadamente en su propio interés. Nuestros intereses tal como los concebimos y los «intereses» de nuestros genes bien pueden diferir, a pesar de que si no fuera por los «intereses» de nuestros genes no existiríamos. Su conservación es nuestra razón de ser original, aun cuando podamos aprender a ignorar ese objetivo e inventar nuestro propio *summum bonum* gracias a la inteligencia, la capacidad de aprender, que nuestros genes han instalado en nosotros. Así, pues, obtenemos nuestra intencionalidad de la de nuestros genes «egoístas». ¡Son ellos los significadores insignificados, no nosotros!

Desde luego, la intencionalidad de nuestros genes no es *intrínseca* en ningún sentido; el «significado» de cualquier gen depende de todo el evolucionado sistema «alfabético» de codones ACGT, síntesis proteica y desarrollo, para empezar. Sin embargo, es *original* en el sentido de que es el *primero* de muchos sistemas *representativos* que han

evolucionado. Todos los sistemas posteriores tienen agentes —sistemas intencionales— cuyas representaciones obtienen su intencionalidad de los objetivos que promueven (exactamente como la intencionalidad del robot gigante).[9]

Esta concepción de las cosas, si bien ofrece una respuesta satisfactoria a la interrogante en torno a de dónde viene nuestra propia intencionalidad, sí parece dejarnos con un bochorno, pues la deriva de entidades —los genes— cuya intencionalidad es un caso paradigmático de mera intencionalidad *como si*. ¿Cómo podría lo literal depender de lo metafórico? Por otra parte, con seguridad mi historia de ciencia ficción y el relato de Dawkins discrepan una de otro en este aspecto: en mi historia supuse que en la creación del robot había participado una ingeniería consciente, deliberada y previsora, mientras que nosotros, aun si somos, como dice Dawkins, producto de un proceso de diseño en el que nuestros genes son los principales beneficiarios, se trata de un proceso de diseño al que le falta completamente un ingeniero consciente, deliberado y previsor. Sin embargo, como veremos, esto no es una buena objeción.

El principal encanto de la teoría de la selección natural es que nos muestra cómo eliminar al artífice inteligente de nuestra explicación sobre los orígenes, y, con todo, el proceso de selección natural es responsable de diseños de enorme ingenio. No son los genes los diseñadores: los genes mismos no podrían ser más tontos; no pueden razonar ni representarse nada ni entender nada. Ellos mismos no hacen el diseño; son tan sólo los *beneficiarios* del proceso de diseño: los clientes, podríamos decir. (En nuestra historia, podrían compararse con un cliente muy tonto y muy rico que contratara a los mejores ingenieros para construirle una máquina de supervivencia. Si no fuera por él, los ingenieros no tendrían un encargo con buen financiamiento, y la supervivencia *del cliente* es lo que paga el artefacto que ellos crean.) ¿Quién o qué hace el diseño? La Madre Naturaleza, desde luego, o más literalmente, el largo y lento proceso de evolución por selección natural.

Para mí, la propiedad más fascinante del proceso evolutivo es su asombrosa capacidad para reflejar *algunas* propiedades de la mente humana (el Diseñador Inteligente) y al mismo tiempo estar desprovisto de otras. Se tratará mucho más de este tema en el capítulo VI, sobre el pensamiento en torno a la evolución, pero mientras tanto quiero aclarar el sólido vínculo que estoy proponiendo entre cualquier teoría aceptable del significado y la teoría de la evolución. Aunque nunca podrá hacerse suficiente hincapié en que la selección natural opera sin previsión ni propósito alguno, tampoco deberíamos perder de vista el hecho de que el proceso de selección natural ha demostrado ser sumamente sensible a la lógica: hace miríadas de «elecciones» con discernimiento, «reconoce» y «aprecia» muchas relaciones sutiles. Para decirlo de manera todavía más provocadora, cuando la selección natural selecciona, puede «elegir» un diseño particular *por una razón más que por otra,* sin jamás «representar»

conscientemente —¡ni inconscientemente!— ya sea la elección o las razones. Los corazones fueron «elegidos» por ser excelentes bombas que hacen circular la sangre, no por el ritmo cautivador de sus latidos, aunque ésta *podría* haber sido la *razón* por la que alguna otra cosa fuera «elegida» por la selección natural.

Tal como el titular de la franquicia panameña de Pepsi-Cola puede seleccionar al tragacuartos *por* su talento al reconocer cuartos de balboa y puede adoptarlo *como* un detector de cuartos de balboa, la evolución puede seleccionar un órgano *por* su capacidad para oxigenar sangre e instituirlo *como* pulmón. Solamente en relación con dichas «elecciones» de diseño o propósitos «aprobados» por la evolución —razones de ser— podemos identificar conductas, acciones, percepciones, creencias o cualquier otra de las categorías de la **psicología vulgar**.[10]

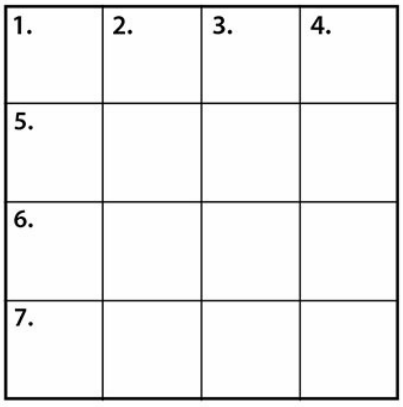
La idea de que somos artefactos diseñados por la selección natural es persuasiva y familiar; algunos dirían tanto como que ni siquiera puede ser tema de una polémica seria.[11] ¿Entonces por qué se resisten tanto a ella no sólo los creacionistas y los ideólogos del «diseño inteligente», sino también (un poco subliminalmente) gente como Searle, Fodor y compañía? Tengo la sensación de que tiene dos implicaciones que no son del todo obvias y que a algunos les parecen terriblemente difíciles aceptar. En primer lugar, si somos «sólo» artefactos, entonces qué significan nuestros pensamientos más íntimos —si es que significan algo a fin de cuentas— es un tema sobre el cual nosotros, los mismísimos pensadores de esos pensamientos, no tenemos mayor autoridad. El tragacuartos se convierte en un detector-cb sin que cambie jamás su naturaleza interna; el estado que solía significar una cosa ahora significa otra. Lo mismo podría pasarnos a nosotros *en principio* si sólo somos artefactos, si por ende nuestra propia intencionalidad no es original sino derivada. (Jones, por ejemplo, no es autoridad para decirnos si está pensando acerca de caballos o chaballos.) En segundo lugar, si somos tales artefactos, no sólo no tenemos la garantía de un acceso privilegiado a verdades más profundas que fijen los significados de nuestros pensamientos, sino que *no existen tales verdades más profundas.* A veces la interpretación funcional es obvia (el corazón es obviamente una bomba; el ojo obviamente sirve para ver), pero cuando no es así, y vamos y leemos la mente de la Madre Naturaleza, no hay texto que interpretar. Cuando la «verdad» sobre la función propia es controvertida —cuando más de una interpretación tiene buen apoyo—, sencillamente no hay tal verdad.

30. Traducción radical y un crucigrama quineano

Fue el filósofo W. V. O. Quine quien defendió de manera memorable y sumamente acertada la afirmación de que cuando dos interpretaciones funcionales igualmente buenas están en conflicto no hay hechos más profundos que pudieran arreglar el asunto.[12] Lo hizo con su principio de la *indeterminación de la traducción radical,* que defendió con ayuda de una famosa bomba de intuición. Imagina que descubres en medio del océano Pacífico, digamos, una isla apartada habitada por gente que no se sabe de dónde vino y que habla un idioma que nadie más conoce. Sin intérpretes bilingües que puedan ayudar, los antropólogos o lingüistas tienen que descifrar el idioma mediante observación e interacciones de ensayo y error con los nativos, una tarea que Quine llamó «traducción radical». Él argumentaba que si a esos investigadores se les diera la tarea de producir a partir de su experiencia un manual de traducción para esta lengua exótica, *en principio* podrían obtener manuales sustancialmente diferentes pero igualmente buenos, que asignaran distintos significados a las emisiones de los nativos, ¡y donde podría no haber modo alguno de saber cuál traducción es la correcta! A muchos filósofos la idea les ha parecido demasiado radical para tomarse en serio, y en consecuencia simplemente la ignoran y persisten con sus estilos tradicionales. Aquí hay una herramienta de pensamiento concebida para que la idea parezca al menos verosímil, si no categóricamente obvia. Es necesario explicar dos cosas: *a)* cómo podría ser verdadera («en principio») la afirmación de Quine, y *b)* cómo, sin embargo, podría ser casi imposible dar un ejemplo real de ella.

Suelo darles este crucigrama a mis alumnos y pedirles que lo resuelvan. A los pocos minutos la mayoría dicen haberlo conseguido. Ahora tú intenta resolverlo antes de seguir leyendo.[13]

FIGURA 10



Horizontales
1. Algo sucio
5. Una gran necesidad humana
6. Hacer suave
7. Actor de cine

Verticales
1. Vehículo que depende de H_2O
2. Normalmente queremos esto
3. Encima
4. Estado de Estados Unidos (abreviado)

¿Lo resolviste? Si así fue, ¿cuál de las soluciones encontraste? El crucigrama tiene dos soluciones, más o menos igual de buenas (están escondidas más adelante en el libro, para darte la oportunidad de encontrar ambas antes de que sean reveladas). Me tomó horas inventarlo, chico como es, porque las múltiples restricciones impuestas interactúan para limitar drásticamente las posibilidades. Si no me crees, ¡trata de hacer uno más grande y mejor! (y si lo consigues, por favor envíamelo; lo usaré en lugar del mío).

Cualquiera que pregunte «¿Cuál es *realmente* la palabra 1 vertical?» sigue condenado a cierta clase de realismo fuera de lugar. *No hay una verdad.* A propósito lo organicé para que no hubiera una verdad. Por ejemplo, no lo hice con un conjunto de respuestas (las respuestas originales o las históricamente primeras y «por lo tanto» las respuestas *reales*), para luego tratar de encontrar otro conjunto. Trabajé juntas las dos soluciones, tomándolas de una lista de parejas de palabras de cuatro letras con parecidos significados que había recolectado.

Es posible hacer un crucigrama así porque hay normas para definiciones que permiten cierta flexibilidad. Las dos soluciones incluyen palabras que apenas encajan en sus definiciones, pero la conspiración de que todo lo que nos rodea encaja (el holismo, en la jerga de los filósofos) mete las palabras en dos configuraciones bastante estables. ¿Cuánto apostarías a que no habrá una tercera solución que compita parejo con cualquiera de estas dos? En general, se cumple la *máxima del criptógrafo:* si puedes encontrar *una* solución a un crucigrama, has encontrado *la única* solución. Sólo circunstancias especiales permiten que haya hasta dos soluciones, pero esos casos nos muestran que la existencia de una sola solución a una pregunta así no es una necesidad metafísica, sino sólo el inmensamente probable resultado de restricciones muy poderosas.

La gente es mucho más complicada que los crucigramas o que las computadoras. Tienen cerebros enrollados llenos de neuromoduladores, que están incorporados a cuerpos profundamente entrelazados con el mundo, y tienen una historia tanto evolutiva como personal que los ha arraigado en el mundo con mucha más interpenetración que el arraigo de un crucigrama en una comunidad lingüística. Entonces Ruth Millikan (por ejemplo) tiene razón en que dada la naturaleza de las restricciones de diseño, es extremadamente poco probable que pudiera haber diferentes maneras de matar pulgas que dejaran dos interpretaciones radicalmente diferentes, globalmente indeterminadas, empatadas en primer lugar. La indeterminación de la traducción radical en la práctica es verdaderamente insignificante. Con todo, el principio sobrevive. La razón por la que no tenemos indeterminación de traducción radical *no es* porque sea un hecho metafísico que haya «verdaderos significados» allá en la cabeza (lo que Quine llamaba el «mito de museo» del significado, su blanco principal). La razón por la que no tenemos indeterminación en el mundo real es que con tantas restricciones independientes que satisfacer, la máxima del criptógrafo nos asegura que es una preocupación

insignificantemente pequeña. Cuando la indeterminación amenaza en el mundo real, siempre son hechos más «conductuales» o «disposicionales» —más de lo mismo— los que nos sacan de apuros para una interpretación determinada, no algún misterioso «poder causal» o una «semanticidad intrínseca». La interpretación intencional casi siempre llega en el último momento a una *sola* interpretación, pero en el catastrófico caso imaginable en el que interpretaciones duales sobrevivieran a todas las pruebas, no habría verdades más *profundas* para determinar cuál fuera la «correcta». Los hechos sí determinan interpretaciones, pero siempre son los hechos «triviales» los que hacen el trabajo.

31. Motores semánticos y motores sintácticos

¿Cómo puede el significado influir en algo? No parece ser la clase de propiedad física, como la temperatura, la masa o la composición química, que pudiera hacer que algo ocurriera. Para lo que los cerebros *sirven* es para extraer *significado* del flujo de energía que incide en sus órganos de los sentidos, a fin de mejorar las perspectivas de los cuerpos que los alojan y les proporcionan su energía. El trabajo de un cerebro es «producir futuro» en forma de previsiones *sobre las cosas del mundo que importan* para dirigir al cuerpo de maneras adecuadas. Los cerebros son órganos muy caros desde el punto de vista de la energía, y si no pueden hacer bien este importante trabajo, no están desquitando su sueldo. En otras palabras, se espera que los cerebros sean *motores semánticos*. De lo que están *hechos* los cerebros son archimillones de piezas moleculares que interactúan de conformidad con las estrictas leyes de la química y la física, en respuesta a formas y fuerzas; en otras palabras, los cerebros de hecho sólo son *motores sintácticos*.

Imagina que vas con los ingenieros y les pides que te construyan un discernidor de billetes de dólar auténticos o, lo que equivale a lo mismo, un detector de billetes falsos: sus especificaciones son que debe poner todos los dólares auténticos en una pila y todos los billetes falsos en la otra. No es posible, dicen los ingenieros: lo que construyamos sólo puede responder a propiedades «sintácticas»: detalles físicos, como el grosor y la composición química del papel, las formas y colores de los estampados de tinta, la presencia o ausencia de otras propiedades físicas difíciles de falsificar. Lo que *pueden* construir, dicen, es un detector de billetes falsos, bastante bueno pero no infalible, basado en tales propiedades «sintácticas». Será caro, pero indirecta e imperfectamente probará los billetes falsos lo bastante bien para desquitar su sueldo.

Cualquier configuración de partes del cerebro está sujeta a las mismas restricciones. Fuerzas fisicoquímicas causarán que haga lo que hace sin importar qué signifique el *input* (o sólo maomeno signifique). No cometas el error de imaginar que los cerebros, por estar vivos, o por estar hechos de proteínas, y no de silicio y metal, pueden detectar directamente los significados gracias a su **tejido maravilloso**. La física siempre triunfará sobre el significado. Un auténtico motor semántico, que responda *directamente* a los significados, es como una máquina de movimiento perpetuo: físicamente imposible. Entonces ¿cómo pueden los cerebros cumplir con la tarea para la que los designaron? Siendo motores sintácticos que pueden *rastrear* o *imitar* la competencia del motor semántico imposible.[14] ¿Pero esto es siquiera posible? Algunos filósofos han argumentado que si la historia microcausal de cómo funciona un cerebro está completa (sin ninguna laguna misteriosa), simplemente no hay lugar para que el significado influya en algo. En el apartado 33 veremos una bomba de intuición que demuestra que esto es falso enseñando cómo las propiedades semánticas —como la verdad, el significado y la

referencia— tienen un papel imprescindible en algunos sencillos procesos causales. Sin embargo, antes de pasar a esa algo compleja bomba de intuición, quiero revisar un modelo simple, que nos permitirá expresar algunas sospechas sobre las bombas de intuición de los filósofos en general y, si todo marcha bien, disipar algunas dudas que pueden estar interfiriendo con la comprensión.

32. El Hombre del Pantano conoce a un tiburón vaca

Se supone que las bombas de intuición deben, de manera limpia y eficiente, bombear la intuición buscada, y luego deben guardarse. Pero un destino común de las bombas de intuición es inspirar un arrebato de refutaciones, contrarrefutaciones, ajustes y ampliaciones. Donald Davidson, uno de los mejores filósofos estadunidenses del siglo XX, una vez me dijo que lamentaba haber inventado esta bomba de intuición, pues provocó un altercado excesivo que si algo iluminaba, lo hacía sólo de manera intermitente. Aquí está el Hombre del Pantano, una de las favoritas de la filosofía, si bien no de Davidson:

> Imagina que cae un rayo sobre un árbol muerto en un pantano; yo estoy parado cerca de ahí. Mi cuerpo se reduce a sus elementos, y por pura coincidencia (y a partir de diferentes moléculas) el árbol se convierte en mi réplica física. Mi réplica, el Hombre del Pantano, se mueve tal como yo lo hacía; de acuerdo con su naturaleza, se va del pantano, encuentra y parece reconocer a mis amigos, y parece devolver sus saludos en inglés. Se muda a mi casa y parece escribir artículos sobre interpretación radical. Nadie nota la diferencia.
>
> Pero *hay* una diferencia. Mi réplica no reconoce a mis amigos; no reconoce nada, empezando porque nunca conoció nada. No puede saber cómo se llaman mis amigos (aunque desde luego parece que sí sabe); no puede recordar mi casa. No puede querer decir lo mismo que yo con la palabra «casa», por ejemplo, pues el sonido «casa» que él hace no fue aprendido en un contexto que le pudiera dar el significado correcto —o significado alguno—. De hecho, no veo cómo puede decirse que mi réplica significara nada con los sonidos que hace, ni que tuviera pensamiento alguno.[15]

No puede haber escapado de la atención de los filósofos que a nuestros colegas de otras áreas —sobre todo en las ciencias— a menudo les cuesta trabajo disimular su diversión incrédula cuando se plantean temas como la Tierra Gemela o el Hombre del Pantano para una consideración aparentemente seria. ¿Están los científicos siendo unos filisteos, delatando sus oídos de artillero frente a las sutilezas de la investigación filosófica, o los filósofos han perdido el contacto con la realidad? **Más bien** (pista) no quisiera decirlo.

Estos extraños ejemplos intentan demostrar una conclusión conceptual u otra reduciendo deliberadamente a cero todas las características no del todo comprendidas de un fenómeno, excepto una, de modo que lo que verdaderamente importa pueda destacar y resultar esclarecedor. El ejemplo de la Tierra Gemela agranda al máximo la similaridad

interna (te despachan a la Tierra Gemela sin darte oportunidad de reparar en este enorme cambio) de modo que pueda demostrarse que el contexto externo es responsable de aquello que nuestras intuiciones nos dicen. La bomba de intuición del Hombre del Pantano mantiene constantes tanto las disposiciones futuras como los estados internos y reduce a cero la «historia». En consecuencia, estos experimentos mentales imitan los experimentos científicos en su diseño, al intentar aislar una interacción determinante entre variables manteniendo constantes otras variables. Un problema con esos experimentos es que la variable dependiente es la intuición —son bombas de intuición— y lo que la imaginación contribuye para la generación de intuiciones es más difícil de controlar de lo que los filósofos han reconocido (varias de las muletas que desarmaremos de hecho inhiben la imaginación de los lectores al distorsionar sus intuiciones y en consecuencia invalidar los «resultados» del experimento mental).

Pero hay otro problema más hondo con esos experimentos. Es juego de niños soñar ejemplos para «demostrar» más conclusiones conceptuales. Supongamos que una vaca diera a luz algo que átomo por átomo fuera indistinguible de un tiburón. ¿Sería un tiburón? Si le plantearas la pregunta a un biólogo, la reacción benévola sería que estás tratando de hacer una broma forzada. O supongamos que un genio maligno pudiera convertir agua en sólido a temperatura ambiente con sólo sonreírle; ¿esa agua del genio sería hielo? Es una hipótesis demasiado tonta para merecer respuesta. Los genios sonrientes, los tiburones vaca, los zombis y los Hombres del Pantano son, según algunos filósofos, lógicamente posibles, aun cuando no son nomológicamente (causalmente) posibles, y también piensan que esto es importante. Yo no. Podemos suponer que la motivación para lanzar una red de contrafactualidad tan grande es que la respuesta que atrapemos nos dirá algo acerca de la *esencia* del tema en cuestión. ¿Pero quién cree en esencias reales de esta clase hoy en día? Yo no.

Podríamos plantear con respecto a los imanes una pregunta paralela tras observar que hubiera varios candidatos a ser el «hacedor de verdad» —la propiedad definitoria o esencia— de los imanes: *a)* todos los imanes son cosas que atraen el acero, y *b)* todos los imanes son cosas que tienen cierta estructura interna (podemos llamarla alineación i). El viejo criterio conductual *a* ¿en algún momento fue remplazado por el nuevo criterio de estructura interna *b,* o este último simplemente explicaba reductivamente aquél? Para descubrirlo, debemos imaginar que les planteamos a científicos las siguientes preguntas tipo Hombre del Pantano. Supongamos que descubres una cosa que atrae el hierro pero no tiene alineación i (como los imanes estándar). ¿La llamarías imán? O supón que descubres una cosa que tiene alineación i pero no atrae el hierro. ¿La llamarías imán? Los físicos responderían que si se enfrentaran a cualquiera de estos objetos imaginarios, tendrían cosas mucho más importantes por las que preocuparse que por cómo llamarlos. Todo su cuadro científico depende de que haya una profunda regularidad entre la

alineación de dipolos atómicos en dominios magnéticos y la atracción del hierro, y el «hecho» de que sea *lógicamente* posible romper esta regularidad tiene para ellos nulo interés. Lo que es interesante, sin embargo, es la real covarianza de factores «estructurales» y «conductuales». Si encuentran alteraciones de las regularidades, ajustan su ciencia de conformidad con ellas y dejan que los términos caigan donde les dé la gana.

¿El Hombre del Pantano tiene pensamientos y habla inglés o no? ¿Un tiburón vaca es un tiburón? Nada como tiburón y se aparea exitosamente con otros tiburones. Oh, ¿pero no te dije? Es indistinguible átomo por átomo de un tiburón, *pero* tiene ADN de vaca en todas las células. ¿Es imposible? No *lógicamente* imposible (dicen los filósofos). Sólo tan obviamente imposible como para volver poco esclarecedora cualquier discusión posterior. Es tan a todas luces físicamente imposible que los «vestigios» de, digamos, los recuerdos de Davidson aparezcan en la estructura del cerebro de un Hombre del Pantano como que un tiburón esté formado de células que contienen ADN de vaca. El Hombre del Pantano puede no ser lógicamente imposible, aunque sólo sea porque las coincidencias cósmicas de la clase que imaginamos que produciría cosas como el Hombre del Pantano son *por definición* no lógicamente imposibles, pero nunca ocurren, así que ¿a quién le importa *lo que diríamos* si pasaran?

«A mí», dice el filósofo que anda a la caza de **preguntas retóricas**. «Creo que es importante siempre definir tus términos con el mayor rigor y abarcar todas las eventualidades lógicamente posibles. Ése es el modo de llegar a la verdad.» ¿Lo es? En el mundo real, la historia pasada y la función futura están unidas por cables multienroscados de evolución, desarrollo y aprendizaje. *Debido a que* el cuerpo de Davidson había tenido la trayectoria particular que durante años tuvo, Davidson tenía todos esos recuerdos, creencias y proyectos, y no hay un sustituto *real* para estos procesos de acumulación naturales. Basar el veredicto en casos imaginarios que contravienen estas condiciones no sirve para nada que yo pueda ver. De hecho, esos elaborados ejemplos me dan la impresión de ser oportunidades manufacturadas para imponer una dicotomía imaginaria para que puedas aplicar el masbienismo impunemente. «No», dice el filósofo, «¡*no* es una falsa dicotomía! *Por mor del argumento* estamos suspendiendo las leyes de la física. ¿No hizo lo mismo Galileo cuando desterró la ficción de su experimento mental?» Sí, pero de la comparación surge una regla general: la utilidad de un experimento mental es inversamente proporcional al tamaño de su desviación de la realidad.

La Tierra Gemela es físicamente imposible, ¡pero no tan imposible como un Hombre del Pantano! (No pienses que la interpretación de los mundos múltiples de la mecánica cuántica, que está ganando aceptación en ciertos ámbitos, muestra que la Tierra Gemela es físicamente posible después de todo. Incluso si hay «allá fuera» un infinito de

universos, entre ellos [¿infinitamente?] muchos que tienen planetas que son casi como la Tierra, no podemos mandar a un terrícola a visitar ninguno de ellos.) En contraste, el viaje del tragacuartos a Panamá no sólo es posible: casos así bien pueden haber pasado. No tenemos que suspender ninguna ley de la naturaleza para imaginarlo con todo el detalle que queramos.

33. Dos cajas negras

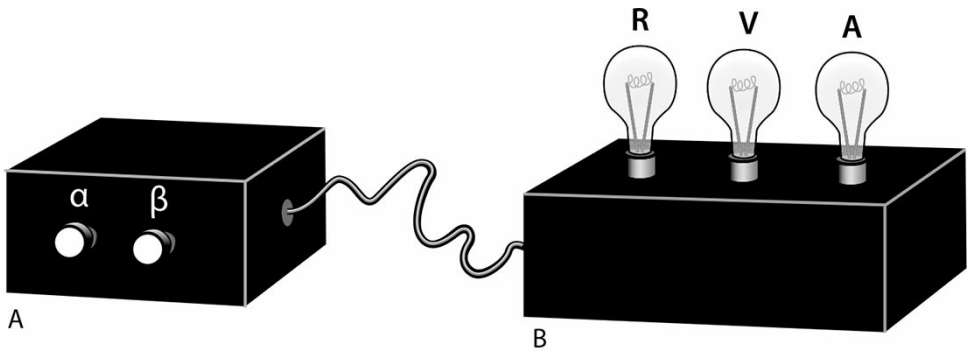


FIGURA 11

Había una vez dos grandes cajas negras, A y B, conectadas con un largo alambre de cobre aislado. En la caja A había dos botones, marcados α y β, y en la caja B había tres luces, roja, verde y amarilla. Unos científicos que estudiaban el comportamiento de las cajas observaron que cada vez que apretaban el botón α de la caja A, la luz roja destellaba brevemente en la caja B, y cada vez que apretaban el botón β de la caja A, la luz verde destellaba brevemente. La luz amarilla nunca parecía destellar. Hicieron unos cuantos miles de millones de ensayos, bajo una muy amplia variedad de condiciones, y no encontraron una sola excepción. Concluyeron que había una regularidad causal, que convenientemente resumieron así:

Todas las α causan rojos
Todas las β causan verdes

Determinaron que de alguna manera la causalidad pasaba a través del alambre de cobre, pues si lo cortaban se apagaban todos los efectos de la caja B, mientras que ocultar las cajas una de otra sin cortar el alambre nunca perturbó la regularidad. Así, pues, naturalmente les daba curiosidad saber exactamente cómo pasaba a través del alambre la regularidad causal que habían descubierto. Tal vez, pensaban, apretar el botón α causaba que se emitiera por el alambre un impulso de bajo voltaje que detonaba la luz roja, y apretar el botón β causaba un impulso de alto voltaje que detonaba la verde. O quizá apretar α causaba un único impulso que detonaba la luz roja, y apretar β causaba un doble impulso. Era claro que debía de haber algo que siempre ocurría en el alambre cuando los científicos apretaban el botón α, y algo diferente que siempre ocurría en el alambre cuando apretaban el β. Descubrir exactamente qué era esto explicaría la regularidad causal que habían descubierto.

Una especie de intervención en el alambre pronto reveló que el asunto era más

complicado. Cada vez que *cualquiera* de los botones se apretaba en la caja A, bajaba velozmente por el alambre a la caja B una larga corriente de impulsos e interrupciones —encendidos y apagados, o bits (10 000 bits, para ser exactos)—. Sin embargo, ¡la configuración de bits era diferente cada vez!

Estaba claro que tenía que haber una característica o propiedad de las cadenas de bits que detonaban la luz roja en un caso y la luz verde en otro. ¿Qué podía ser? Los científicos decidieron abrir la caja B y ver qué pasaba con las cadenas de bits cuando llegaban. Adentro de B los científicos encontraron una supercomputadora en serie digital común y corriente, con una gran memoria, que contenía un enorme programa y una enorme base de datos, escritos, desde luego, en más cadenas de bits. Y cuando rastrearon los efectos de las cadenas de bits entrantes en este programa de computadora, no encontraron nada extraordinario: la cadena de entrada siempre se abría paso hacia el CPU (siglas de *unidad de procesamiento central* en inglés) de la manera habitual, donde provocaría que se realizaran unos cuantos miles de millones de operaciones en pocos segundos, para siempre terminar con una de dos señales de salida, un 1 (que encendía la luz roja) o un 0 (que encendía la luz verde). Descubrieron que siempre podían explicar en el nivel microscópico cada paso de la causalidad sin ninguna dificultad o polémica. No sospechaban que operara ninguna causa oculta, y, por ejemplo, cuando dispusieron que entrara una y otra vez la misma secuencia de 10 000 bits, el programa de la caja B siempre tenía el mismo resultado de salida, rojo o verde.

Esto, empero, era ligeramente desconcertante, porque si bien la caja B siempre tenía un mismo resultado de salida, no lo hacía por pasar a través de los mismos pasos intermedios. De hecho, casi siempre pasaba por distintos estados físicos antes de arrojar la misma salida. Esto de suyo no era ningún misterio, pues el programa guardaba una copia de cada entrada que recibía, y así, cuando la misma entrada llegaba por segunda, tercera o milésima vez, el estado de la memoria de la computadora era ligeramente distinto cada vez. Sin embargo, la salida era siempre la misma; si la luz se ponía roja la primera vez que entraba una cadena de bits particular, en lo sucesivo siempre se ponía roja con esa misma cadena, y se cumplía la misma regularidad con las cadenas verdes (como los científicos empezaron a llamarlas). Estaban tentados de proponer la hipótesis de que todas las cadenas son o bien rojas (causan un destello de luz roja) o verdes (causan un destello de luz verde), pero desde luego que no probaron con todas las cadenas posibles: sólo con cadenas que la caja A había emitido.

Los científicos decidieron probar sus hipótesis desconectando temporalmente A de B e insertando variaciones en las cadenas de salida de A a B. Para su gran consternación, descubrieron que casi siempre, cuando alteraban una cadena de A, ¡destellaba la luz *amarilla*! Era casi como si la caja B hubiera detectado su intervención. No había duda, sin embargo, de que la caja B siempre aceptaría versiones de cadenas rojas hechas por

personas, y destellaría rojo en respuesta, y aceptaría cadenas verdes hechas por personas, y destellaría verde en respuesta. Sólo cuando se cambiaba un bit —o más de uno— en una cadena roja o verde, normalmente —casi siempre— se encendía la luz amarilla. «¡La mataste!», espetó alguien después de ver cómo una cadena «alterada» se convertía en cadena amarilla, y esto provocó un aluvión de especulaciones sobre la posibilidad de que las cadenas rojas y verdes estuvieran *vivas* en algún sentido —y que quizá fueran macho y hembra—, y que las amarillas fueran cadenas *muertas*. Esta hipótesis, atractiva como era, resultó no llevar a ningún lado, aunque un aluvión de experimentos con unos cuantos miles de millones de variaciones al azar de cadenas de 10 000 bits sugería convincentemente a los científicos que realmente había tres variedades de cadenas —rojas, verdes y amarillas— y que las amarillas superaban en número a las rojas y las verdes por muchísimos órdenes de magnitud (véase el apartado 35 para más sobre esto). Casi todas las cadenas eran amarillas. Eso hacía todavía más emocionante y desconcertante la regularidad rojo/verde que habían descubierto.

¿Qué tenían las cadenas rojas que encendía la luz roja y las cadenas verdes que encendía la luz verde? *Por supuesto que en cada caso particular no había ningún misterio.* Los científicos podían rastrear la causalidad de cada cadena particular a través de la supercomputadora de B y ver que, con un determinismo gratificante, producía su luz roja, verde o amarilla, según el caso. Lo que no pudieron encontrar, empero, fue una manera de predecir cuál de los tres efectos tendría una nueva cadena con tan sólo revisarla (sin «simular a mano» su efecto en la caja B). Por sus datos empíricos sabían que había muy altas probabilidades de que cualquier nueva cadena que se estudiara sería amarilla, a menos que se tratara de una cadena que supieran que había sido emitida por la caja A, en cuyo caso las probabilidades eran mayores que mil millones a uno de que sería *o* roja *o* verde, pero nadie podía saber cuál sin ejecutarla en la caja B para ver cómo lo resolvía el programa.

A lo mejor la solución al misterio estaba en la caja A. La abrieron y encontaron otra supercomputadora, de diferente marca y modelo, y que ejecutaba otro programa enorme, pero también se trataba de una computadora digital de lo más común y corriente. Pronto determinaron que había un «reloj» interno haciendo tictac millones de veces por segundo, y cada vez que apretaban cualquier botón, lo primero que hacía la computadora era tomar la «hora» del reloj (por ejemplo, 101101010101010111) y descomponerlo en cadenas, que luego usaba para determinar qué subrutinas activar en qué orden, y a qué parte de su memoria acceder primero en el curso de su preparación de una cadena de bits para mandar por el alambre.

Los científicos pudieron darse cuenta de que era esta consulta del reloj (que es prácticamente aleatoria) lo que virtualmente garantizaba que nunca se mandara dos veces la misma cadena de bits. Sin embargo, a pesar de esta aleatoriedad, o seudoaleatoriedad,

seguía siendo cierto que cada vez que apretaban el botón α, la cadena de bits que la computadora inventaba resultaba ser roja, y cada vez que apretaban el botón β, la cadena de bits que tarde o temprano mandaba resultaba ser verde. De hecho, los científicos sí encontraron algunos casos anómalos: aproximadamente en un ensayo por cada mil millones, apretar el botón α causaba que se emitiera una cadena verde, o apretar el botón β causaba que se emitiera una cadena roja. Esta pequeña imperfección sólo consiguió avivar las ansias de los científicos por encontrar una explicación de la regularidad.

Un buen día llegaron los dos *hackers* expertos en inteligencia artificial que habían construido las cajas, y explicaron todo (no sigas leyendo si quieres entender el misterio por ti mismo). Al, que había construido la caja A, había estado trabajando años en un «sistema experto» —una base de datos que contenía «proposiciones verdaderas» acerca de todas las cosas, y un motor de inferencias para deducir otras implicaciones de los axiomas que componían la base de datos—. En la base de datos había estadísticas de beisbol de las Ligas Mayores, registros meteorológicos, taxonomías biológicas, historias de los países del mundo y gran cantidad de banalidades. Bo, el sueco que había construido la caja B, al mismo tiempo había estado trabajando en una base de datos rival sobre «conocimiento del mundo» para su propio sistema experto. Cada uno llenó su respectiva base de datos con cuantas «verdades» le habían permitido los años de trabajo.[16]

Sin embargo, con el paso de los años se habían aburrido de los sistemas expertos y los dos decidieron que la promesa práctica de esta tecnología estaba muy sobrestimada. De hecho, los sistemas no eran tan buenos resolviendo problemas interesantes ni «pensando» ni «encontrando soluciones creativas a los problemas». Lo único para lo que eran buenos, gracias a sus motores de inferencia, era para generar cantidades enormes de oraciones verdaderas (en sus idiomas respectivos) y probar la verdad y falsedad de oraciones que les suministraran como entrada (en sus idiomas respectivos), en relación con su *maomeno* conocimiento. Así que Al y Bo se habían juntado y resuelto cómo podrían emplear los frutos de su esfuerzo desperdiciado. Decidieron hacer un juguete filosófico. Eligieron una lengua franca para que se tradujeran sus dos sistemas representacionales (de hecho fue el español, y se mandaba en código ASCII estándar), y engancharon las máquinas con un alambre.[17] Cada vez que apretabas el botón α de A, se le daba a A la instrucción de elegir aleatoriamente (o seudoaleatoriamente) una de sus «creencias» (ya fuera un axioma almacenado o una implicación generada de sus axiomas), traducirla al español (en una computadora, los caracteres españoles ya estarían en ASCII), agregar suficientes bits aleatorios después del punto para que sumaran 10 000, y mandar la cadena resultante a B, que traducía esta entrada a su propio idioma (que era el Lisp sueco), y la comparaba con sus propias «creencias» —su base de datos—. Como las dos bases de datos estaban hechas de verdades, y aproximadamente las mismas

verdades, gracias a sus motores de inferencia, cada vez que A mandaba a B algo que A «creía», B también lo «creía», y lo señalaba destellando una luz roja. Cada vez que A mandaba a B algo que A consideraba una falsedad, B destellaba una luz verde en señal de que le parecía que en efecto era una falsedad.

Y cada vez que alguien alteraba la transmisión, esto casi siempre traía como resultado una cadena que no era un enunciado en español bien construido, a menos que la alteración ocurriera solamente en la basura aleatoria del final. Como B no tenía absolutamente ninguna tolerancia para los errores tipográficos, B respondía a ellos destellando la luz amarilla. Cada vez que alguien elegía aleatoriamente una cadena de bits, había una enorme probabilidad de que no fuera una verdad o falsedad bien construida en ASCII español, de ahí el predominio de las cadenas amarillas.

Entonces, dijeron Al y Bo, la misteriosa propiedad causal *rojo* era en realidad la propiedad de ser una oración verdadera en español, y *verde* era la propiedad de ser una falsedad en español. De pronto, la búsqueda que durante años había eludido a los científicos se había convertido en un juego de niños. Cualquiera podía componer cadenas rojas *ad nauseam.* Bastaba con escribir el código ASCII para «Las casas son más grandes que los cacahuates», «Las ballenas no vuelan» o «Tres por cuatro es dos menos que dos por siete», por ejemplo. Si querías una cadena verde, podías probar con algo como «Nueve es menor que ocho» o «Nueva York es la capital de España». Pronto a los filósofos se les ocurrieron algunos trucos simpáticos, como encontrar cadenas que fueran rojas las primeras 100 veces que se les dieran a B pero verdes en lo sucesivo (por ejemplo, el código ASCII por «Se te ha enviado este enunciado para evaluación menos de 101 veces»).

Sin embargo, dijeron algunos filósofos, las propiedades de cadenas *roja* y *verde* en realidad no son *verdad en español* y *falsedad en español.* Después de todo, hay verdades en español cuyas expresiones en ASCII tienen millones de bits, y además, a pesar de sus mayores esfuerzos, Al y Bo no siempre insertaban *hechos* en sus programas. Por ejemplo, algo que pasaba por vox pópuli cuando estaban trabajando en sus bases de datos, para entonces ya había sido desmentido. Había muchas razones por las que la propiedad de cadena —la propiedad causal— de *rojez* no era exactamente la propiedad de *verdad en español.* Así, a lo mejor *rojo* podía quedar mejor definido como una *expresión relativamente corta en español ASCII de algo que la caja B (cuyas maomeno creencias casi todas son verdaderas)* **maomeno** *cree que es verdadera.* Esto satisfizo a algunos, pero otros le encontraron defectos e insistieron en que, por varias razones, esta definición era inexacta, o tenía contraejemplos que no podrían descartarse de ninguna manera no *ad hoc,* etc. Sin embargo, como Al y Bo señalaron, no podían encontrarse mejores descripciones de la propiedad, y ¿no habían estado anhelando los científicos precisamente una explicación así? ¿No se había resuelto ya por completo el

misterio de las cadenas rojas y verdes? Es más, ahora que estaba resuelto, ¿no podía uno ver que no había absolutamente ninguna esperanza de explicar la regularidad *causal* con la que empezó este cuento —todas las α causan rojos y todas las β causan verdes— sin usar *algunos* términos *semánticos* (o *mentalistas*)?

Algunos filósofos argumentaron que si bien la recién descubierta descripción de la regularidad en la actividad del alambre podía usarse para predecir el comportamiento de la caja B, a fin de cuentas no era una regularidad *causal*. La verdad y la falsedad (y cualquiera de los suplentes ajustados que acaban de considerarse) son propiedades semánticas y, como tales, son enteramente abstractas, y por lo tanto no podrían *causar* nada. Tonterías, replicaban otros. Apretar el botón α causa que se encienda la luz roja con tanta certeza como dar la vuelta a la llave de encendido causa que tu auto arranque. Si hubiera resultado que lo que estaba mandándose por el alambre era simplemente alto voltaje contra bajo voltaje, o un impulso contra dos, todo el mundo habría estado de acuerdo en que éste era un sistema causal ejemplar. El hecho de que resultara ser una máquina de Rube Goldberg no mostraba en lo absoluto que la fiabilidad del vínculo entre α y los destellos rojos no fuera causal. De hecho, en todos y cada uno de los casos, los científicos pudieron rastrear el preciso camino microcausal que explicaba el resultado.[18]

Convencidos de esta línea de razonamiento, otros filósofos empezaron a argumentar que esto mostraba que a fin de cuentas las propiedades rojo, verde y amarillo no eran *realmente* semánticas o mentalistas, sino sólo propiedades semánticas de imitación, simples propiedades semánticas «como si». Lo que realmente eran el verde y el rojo eran propiedades *sintácticas* muy pero muy complicadas. Sin embargo, estos filósofos declinaron decir nada más sobre exactamente qué propiedades sintácticas eran éstas, o explicar cómo hasta los niños pequeños podrían rápida y confiablemente producir ejemplos de ellas o reconocerlas. Estos filósofos estaban, no obstante, convencidos de que *tenía* que haber una descripción puramente sintáctica de la regularidad, pues, después de todo, los sistemas causales en cuestión eran «sólo» computadoras, y las computadoras son «sólo» motores sintácticos, incapaces de tener ninguna «semanticidad» real.

«Suponemos», replicaron Al y Bo, «que si nos hubieran encontrado *a nosotros* dentro de nuestras cajas negras, siguiendo el mismo esquema por hacerles una broma, transigirían y estarían de acuerdo en que la propiedad causal operativa era genuinamente verdadera (o creída verdadera, en todo caso). ¿Pueden proponer alguna buena razón para hacer una distinción así?» Esto hizo que algunos declararan que en un sentido importante Al y Bo *habían* estado en las cajas, pues eran responsables de haber creado las bases de datos respectivas como modelos de sus propias creencias. A otros les hizo negar que realmente hubiera propiedades semánticas o mentalistas en ningún lugar del mundo. El contenido, decían ellos, se había *eliminado*. El debate prosiguió durante años,

pero el misterio con el que comenzamos estaba resuelto.

Bloquear las salidas

Aquí termina el cuento de las dos cajas negras. La experiencia muestra, sin embargo, que no es posible presentar un experimento mental de forma tan clara que ningún filósofo pueda malinterpretarlo, así que a fin de prevenir algunas de las malas interpretaciones más atractivas, con total falta de elegancia llamaré la atención sobre algunos de los detalles críticos y explicaré los papeles que representan en esta bomba de intuición.

1. Los dispositivos de las cajas A y B no son más que enciclopedias automatizadas; ni siquiera «enciclopedias ambulantes» sino sólo «cajas de verdades». Ningún aspecto de la historia presupone o implica que estos dispositivos sean conscientes, o sean *cosas pensantes,* o siquiera *agentes,* salvo en el sentido mínimo en que un termostato es un agente. Son sistemas intencionales completamente aburridos, rígidamente arreglados para cumplir un solo y sencillo objetivo (lo mismo puede decirse de Watson de IBM, desde luego). Contienen grandes cantidades de proposiciones verdaderas y de la maquinaria inferencial necesaria para generar más verdades y para probar la «verdad» comparando una proposición candidata con sus bases de datos existentes.
2. Como los dos sistemas fueron creados de manera independiente, no es verosímil que debieran contener (ni real ni virtualmente) *exactamente* las mismas verdades, pero para que la broma del cuento funcionara tan bien como digo, debemos suponer una grandísima coincidencia, de modo que fuera extremadamente poco probable que A generara una verdad que B no reconociera como tal. Sostengo que dos consideraciones hacen que esto sea convincente: *a)* Al y Bo pueden vivir en diferentes países y tener distintas lenguas maternas, pero *viven en el mismo mundo,* y *b)* aunque hay archimillones de proposiciones verdaderas sobre ese mundo (nuestro mundo), el hecho de que Al y Bo se hayan propuesto crear bases de datos *útiles* garantizaría una gran coincidencia entre los dos sistemas de creación independiente. Aunque Al podría saber que cuando cumplió veinte años su pie izquierdo estaba más cerca del Polo Norte que del Polo Sur, y Bo no había olvidado que su primer maestro de francés se apellidaba Dupont, con toda seguridad éstas no serían verdades que ninguno de ellos pondría en su respectiva base de datos. Si dudas que el mero hecho de que ambos estuvieran concentrados en crear una enciclopedia útil en todo el mundo garantizaría tan cercana correspondencia entre sus bases de datos, sólo añade, como un detalle poco elegante, el hecho conveniente de que durante los años que fueron *hackers* intercambiaron impresiones sobre los temas que debían ser cubiertos.
3. ¿Por qué no poner en el cuento a Al y Bob (un compañero estadunidense) o,

si a ésas vamos, por qué no poner simplemente un duplicado del sistema de Al en la caja B? Porque es de fundamental importancia para mi cuento que no haya una correspondencia *sintáctica* simple y factible de ser descubierta que pudiera explicar la regularidad. Por eso el sistema de Bo está en Lisp sueco —para esconder de miradas indiscretas las subyacentes características *semánticas* compartidas entre las estructuras de datos consultadas mientras A se daba a la tarea de generar enunciados y B a la de traducir enunciados y confirmar su verdad—. Las computadoras, como sistemas físicos, deben ser, en el mejor de los casos, motores sintácticos que responden directamente a diferencias físicamente transducibles, no a significados. Sin embargo, tanto A como B se han diseñado para reflejar con la mayor exactitud posible el sabelotodo imaginario, un motor *semántico* lleno de verdades *entendidas.* Cuando dos sistemas sintácticos *diferentes,* A y B, se han diseñado para reflejar el *mismo* motor semántico, la única manera de explicar la notable regularidad que muestran es ascender al nivel de motor semántico, donde las verdades se creen y se pretende hacer aseveraciones. La idea, pues, era crear dos sistemas que presentan la fascinante regularidad del comportamiento externo descrito pero que internamente fueran lo más diferentes posible, de modo que *sólo* el hecho de que sus respectivas tripas fueran *representaciones sistemáticas de un mundo común* pudiera explicar la regularidad (recuérdese que éste era el tema del apartado 13).

Podríamos hacer una pausa para preguntar si dos sistemas así podrían o no ser tan inescrutables como para que la ingeniería inversa no los afectara. En otras palabras, ¿podrían los científicos haberse quedado perplejos tanto tiempo? La criptografía se ha movido a regiones tan arcanas y enrarecidas que uno debería pensárselo al menos tres veces antes de pronunciarse por cualquiera de estas posibilidades. No tengo idea de si alguien puede elaborar un argumento razonable de que hay esquemas de cifrado irrompibles o de que no los hay. Con todo, más allá del cifrado, los *hackers* apreciarán que todos los comentarios y otros indicadores que uno pone en el código fuente cuando escribe un programa se desvanecen cuando se compila el código fuente, dejando atrás una maraña de instrucciones para la máquina *casi* imposible de descifrar. «Descompilar» —hacer ingeniería inversa del código objeto y recuperar el código fuente— a veces es posible en la práctica (¿es siempre posible en principio?), aunque no restablece los comentarios sino que sólo hace que destaquen las estructuras en el lenguaje de más alto nivel. Mi suposición de que los empeños de los científicos para descompilar el programa y descifrar las bases de datos se malogró podría reforzarse postulando el cifrado, si hiciera falta. En el relato tal como lo he contado, podemos estar de acuerdo en que es extraño que a los científicos nunca se les hubiera ocurrido revisar si había alguna traducción a ASCII de las cadenas de bits que corrían a través del alambre. ¿Cómo

pudieron ser tan burros? Está bien: puedes arreglar esta falla en el experimento mental si mandamos todo el artilugio (las cajas A y B y el alambre que las conecta) a «Marte», y dejamos que los científicos extraterrestres traten de entender la regularidad. El hecho de que todas las α causaran luces rojas, todas las β causaran luces verdes y las cadenas de bits aleatorios causaran luces amarillas sería tan visible para ellos como para nosotros, pero no van a tener la menor idea sobre ASCII. Para ellos, este regalo del espacio exterior presentará una regularidad completamente misteriosa, totalmente por encima de cualquier investigación analítica *hasta* que se les ocurra la idea de que cada caja contiene una *descripción de un mundo,* y que las descripciones son *del mismo mundo.*[19] Es el hecho de que cada caja mantiene relaciones semánticas variopintas con las mismas cosas, si bien expresadas en «terminología» diferente y axiomatizadas de distinto modo, lo que fundamenta la regularidad.

Cuando probé este experimento mental con Danny Hillis, creador de la Máquina de Conexión, una computadora masivamente paralela pionera que a principios de la década de 1980 construyó su compañía, Thinking Machines, inmediatamente pensó en una «solución» criptográfica al enigma, y luego concedió que mi solución podía verse provechosamente como un ejemplo especial de su solución: «¡Al y Bo estaban usando *el mundo* como "libreta de un solo uso"!» —una acertada alusión a una técnica estándar de encriptamiento—. Puedes ver de qué se trata imaginando una variación. Tu mejor amigo y tú están a punto de ser capturados por fuerzas hostiles (piratas del espacio, digamos) que pueden saber inglés pero no mucho acerca de tu mundo. Los dos conocen el código Morse y dan con el siguiente esquema de encriptamiento improvisado: si dices una oración verdadera vale como raya; si dices una oración falsa, vale como punto. A sus captores se les permite oírlos hablar: «Los pájaros ponen huevos, y los sapos vuelan. Chicago es una ciudad, y mis pies no están hechos de hojalata, y el beisbol se juega en agosto», dices, y respondes «no» (raya-punto; raya-raya-raya) a cualquier cosa que tu amigo haya preguntado. La siguiente vez que tengas que decir que no, usas diferentes oraciones. Aunque sus captores conozcan el código Morse, a menos que pudieran determinar también la verdad y falsedad de estas oraciones no podrían detectar las propiedades que representan punto y raya. Para darle sabor, podría añadirse esta variación a nuestra fábula, así: en vez de mandar los sistemas informáticos a Marte, ponemos en las cajas a Al y Bo y los mandamos a ellos. Los marcianos estarán tan desconcertados con Al y Bo, si juegan la broma del código Morse, como con las computadoras, a menos que saquen la conclusión (obvia para nosotros pero no para los marcianos) de que esas cosas en las cajas deben interpretarse semánticamente. De lo que trata la fábula es sencillo. No hay sustituto de la postura intencional: o la adoptas, y para explicar las sucesiones repetidas de puntos y rayas encuentras los mismos hechos de nivel semántico, o pasarás el resto de tu vida perplejo con la regularidad —la regularidad

causal— que manifiestamente está ahí.[20]

En este momento, si eres como muchos filósofos, puede ser que otra vez te atraiga la afirmación de que esta bomba de intuición «funciona» sólo porque las cajas A y B son artefactos cuya intencionalidad, tal como es, es completamente derivada y, digamos, *artefáctica.* Las estructuras de datos en sus memorias obtienen su referencia (si es que obtienen alguna) de la confianza indirecta en los órganos sensoriales, las historias de vida y los propósitos de sus creadores, Al y Bo. La verdadera fuente del significado, la verdad o la semanticidad de los artefactos radica en estos artífices humanos. Al y Bo tienen intencionalidad original, y A y B tienen sólo intencionalidad derivada (desde luego, a eso apuntaba la sugerencia de que en cierto sentido Al y Bo *estaban* en sus cajas respectivas). *Podría* haber contado la historia de otra manera: adentro de las cajas había dos *robots,* Al y Bo, y cada uno había pasado toda una «vida» recorriendo el mundo y recolectando hechos antes de meterse en sus cajas respectivas. Elegí una ruta más simple, para evitar todas las preguntas de si la caja A o la caja B «realmente estaba pensando», pero si quieres reconsiderar el experimento mental con esta complicación, observa que la bomba de intuición acerca de la máquina de supervivencia que es el robot gigante ya ha despertado dudas sobre la idea, por lo demás atractiva, de que no podría surgir intencionalidad genuina en ningún artefacto.

Resumen

Veintiuna herramientas de pensamiento —una docena de bombas de intuición y algunos conceptos útiles— formadas y puestas a trabajar en el concepto fundacional de significado. ¿Qué hemos hecho con su ayuda? ¿Todas funcionan bien? Observa que hay dos maneras en que una bomba de intuición puede resultar valiosa. Si está bien hecha, entonces o bien las intuiciones que bombea son fiables y convincentes, en cuyo caso bloquea muy bien algún camino de error que por lo demás podría ser tentador, o bien las intuiciones *siguen* pareciendo discutibles, en cuyo caso la bomba de intuición puede ayudar a concentrar la atención en lo que anda mal con sus propias presuposiciones. En cualquiera de los dos casos, la herramienta es una especie de palanca, como un subibaja —si un lado sube, el otro lado tiene que bajar—, pero sólo si la herramienta no puede doblarse o romperse a la mitad. Las bombas de intuición son más complicadas que los subibajas, así que tenemos que girar las perillas. Eso es lo que acabo de hacer con las dos cajas negras, pero sin duda hay otras perillas que revisar antes de conformarse con su veredicto.

¿Y cuáles son los veredictos de esta larga sección sobre el significado? Una mezcla de pesimismo y optimismo. El significado no va a resultar ser una simple propiedad que fácilmente encuentre una correspondencia en los cerebros, y en ningún lugar vamos a encontrar verdades «más profundas» que simplemente *resuelvan* la pregunta de qué *significa realmente* una oración, un pensamiento o una creencia. A lo mucho —y esto ya es bastante bueno— podemos encontrar y anclar (aparentemente) *mejores interpretaciones* de todos los datos de la postura física (y la postura de diseño) que tengamos. Si podemos encontrar *una* solución a un dilema quineano acerca del significado, es casi seguro que habremos encontrado *la* solución, en el sentido de que podemos tener la seguridad de que no hay ninguna mejor solución sin descubrir. Como mostró la mudanza del tragacuartos a Panamá, el significado siempre es relativo a la función que se cumpla, y no hace falta que haya ninguna intencionalidad *original* más allá de la intencionalidad de nuestros genes maomeno egoístas, que obtienen su *maomeno* intencionalidad del contexto funcional de la evolución por selección natural, y no de un Diseñador Inteligente que represente el papel de nuestro cliente rico que encarga un robot gigante. Y la bomba de intuición acerca de las **dos cajas negras** muestra que la **postura intencional**, con toda su tolerancia a las **maomeno** creencias y demás calaña, *no es opcional* si quieres entender y explicar muchas de las regularidades causales del mundo.

Ahora en el siglo XXI, nuestras investigaciones sobre estos temas se han enriquecido enormemente por el hecho de que, por primera vez, tenemos maneras de pensar rigurosa y vigorosamente sobre mecanismos con *billones* de partes móviles: partes que funcionan de maneras no misteriosas. Gracias a Turing, ahora podemos ver al menos borrosamente un camino que va de la materia bruta que carece de entendimiento, a través de una

cascada de reorganizaciones (la postura de diseño y los *maomeno* significados), a una apreciación de nosotros mismos como creyentes, conocedores y comprendedores paradigmáticos (que la postura intencional simplifica como sistemas intencionales).

Todas estas proposiciones son, o han sido, polémicas, y todavía hay muchos expertos que no se han afiliado a todas. A lo mejor ponerlas en fila aumentará su poder de convencimiento y obtendremos un fenómeno de «masa crítica» que atraiga a quienes no han visto lo bien que combinan las partes. Por otro lado, si se ponen en fila puede ser más fácil que los críticos encuentren una misma hebra de error que las atraviese todas. De cualquier manera habremos progresado. O quizá descubramos prestidigitación en muchos de ellos, lo que oscurecerá la verdad en vez de iluminarla. Encontrar fallos, regresar y girar un poco más las perillas, y ver qué pasa; como mínimo, descubrir esas fallas equivale a un progreso un poco amargo: el de poner en evidencia malas ideas tentadoras, algo que durante milenios han estado haciendo los filósofos.

No digo que esta sección nos haya dado una *teoría del significado*. Nos da únicamente un espacio lógico algo amplio en el que debe caber una teoría científica del significado apropiada, si estoy en lo correcto.[21] Ahora que hemos precisado el lado de la mente correspondiente al *contenido,* ¿podemos pasar a ese gran misterio de misterios, la conciencia? Todavía no. Tenemos que construir algunos otros cimientos. En nuestro intento de entender la conciencia surgen demasiados temas que tienen implicaciones o presuposiciones sobre la evolución, como vimos en este apartado. Muchos de los temas que hemos empezado a desarrollar han apelado a consideraciones evolutivas, así que expongámoslos y aclarémoslos antes de seguir adelante. Aparte de todo es un tema que nunca deja de ser fascinante.

[1] Una acotación para los expertos: cyc (del inglés *encyclopedia;* D. B. Lenat y R. V. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the cyc Project,* Addison-Wesley, Reading, 1990) es sin duda lo más impresionante que la inteligencia artificial ha puesto en práctica de algo parecido a un lenguaje del pensamiento: una base de datos enciclopédica, en gran medida codificada «a mano», presidida por un motor de inferencia patentado. cyc ha estado desarrollándose por más de un cuarto de siglo ya y muchas manos han trabajado en él, pero obtiene su poder del hecho de tener un diseño despiadadamente no biológico y no psicológico (véase la excelente entrada de Wikipedia en inglés sobre cyc). Es casi seguro que el algo de Mike con los pelirrojos no es una microteoría sobre los pelirrojos axiomatizada compuesta en cerebrés e incorporada a una gran base de datos como cyc. Todavía no sabemos cuánto pueden hacer una gran cantidad de cosas acerca de cosas de este tipo porque todavía no las hemos estudiado directamente, excepto en modelos muy simples (como las arquitecturas de subsunción insectoide de Rodney Brooks y sus colegas —véase R. A. Brooks, *Planning is Just a Way of Avoiding Figuring out What to Do Next,* MIT Artificial Intelligence Laboratory, Cambridge, 1987). Uno de los principales intereses teóricos de su proyecto Cog (un robot humanoide —véase D. C. Dennett, «The Practical Requirements for Making a Conscious Robot», *Proceedings of the Royal Society, A,* 349: 133-146, 1994—) era que llevaba estos modelos profundamente no proposicionales de estructuras portadoras de contenido a un territorio que se puede reconocer como psicología humana.

[2] J. Searle, «Minds, Brains and Programs», *op. cit.*

[3] [Aunque en esta versión en español se ha adaptado «two-bitser» como «tragacuartos», se ha considerado de interés para el lector conservar la siguiente nota de Dennet. (t.)] Probablemente no debí haber usado este terminajo obsoleto para un detector de monedas de 25 centavos de dólar cuando creé esta bomba de intuición, pero lo hice, y el término ha tenido, sí, una gran circulación, así que quedémonos con él. El oscuro origen del término «two bits» para referirse a una moneda de 25 centavos de dólar se remonta al «real de a ocho», a los doblones y a otras reliquias de tiempos de los piratas.

[4] R. Millikan, *Language, Thought and Other Biological Categories,* The MIT Press, Cambridge, 1984.

[5] Una plancha de hierro —como la que tu abuela usaba para planchar la ropa de la familia— puede ser un elegante tope para puertas. Si buscas una en tiendas de antigüedades y sitios web de «objetos coleccionables», asegúrate de que te den una auténtica, no una réplica. Algunas que están a la venta son sólo topes de hierro para puerta fundidos con forma de planchas antiguas. Dentro de cien años alguien podrá manufacturar curiosos topes de puerta que parezcan tragacuartos: ya sabes, los dispositivos que tus abuelos usaban cuando el dinero circulaba en forma de monedas.

[6] R. Millikan, *op. cit.*

[7] El filósofo Hilary Putnam inventó hace muchos años la Tierra Gemela, y mi bomba de intuición ha sido cuidadosamente modelada para reproducir los detalles de la suya que vienen a cuento. Mi historia del tragacuartos es en realidad un ajuste muy elaborado de las perillas de su bomba de intuición. En los últimos 35 años los filósofos han discutido decenas o acaso cientos de otras variantes. En su formulación original, Putnam eligió agua en la Tierra y sucedáneo de agua (compuesta de XYZ, no H_2O) en la Tierra Gemela, pero esto plantea complicaciones que no son vienen al caso en nuestro uso de la bomba de intuición, así que dejemos que en esta historia los chaballos sean los análogos de los cuartos de balboa. Véase H. Putnam, «The Meaning of "Meaning"», en K . Gunderson (comp.), *Language, Mind and Knowledge,* University of Minnesota Press, Minneapolis, 1975.

[8] R. Dawkins, *The Selfish Gene,* Oxford University Press, Oxford, 1976.

[9] El importante artículo de T. Fitch «Nano-Intentionality: A Defense of Intrinsic Intentionality» (*Biology and Philosophy,* 23 [2]: 157-177, 2008) propone que las células eucariotas —pero no sus ancestros, las células procariotas— fueron las primeras entidades evolucionadas con intencionalidad «intrínseca», porque, según afirma, sus talentos autoprotectores eran sorprendentemente más avanzados que los de sus ancestros. Su hincapié en la autonomía, el albedrío, de las células individuales tuvo una influencia considerable sobre mi

funcionalismo homuncular (véase el apartado 20), pero estoy en desacuerdo con su intento de detener en los eucariontes la regresión finita. No es más fácil «remplazar con una máquina» un procarionte que un eucarionte; la agencia baja completamente hasta las proteínas y toca fondo en los genes egoístas.

10 Para un categórico y apasionado rechazo de todo lo dicho en este párrafo, véase J. Fodor y M. PiatelliPalmarini, *What Darwin Got Wrong ,* Farrar, Straus & Giroux, Nueva York, 2010. O bien el libro que estás leyendo está totalmente equivocado, o bien aquel otro lo está. Para una defensa más detallada de la postura de este libro, véanse el capítulo vi y mi artículo «The Evolution of Reasons», en B. Bashour y H. D. Muller (comps.), *Contemporary Philosophical Naturalism and its Implications,* Routledge, Nueva York, 2014, pp. 47-62.

11 Ruth Millikan ha abundado en esta afirmación con mucho mayor detalle que yo en una serie de libros brillantes a partir de 1984. Véase D. Ryder, J. Kingsbury y K. Williford (comps.), *Millikan and Her Critics,* WileyBlackwell, Oxford, 2013, para el estado actual de la cuestión.

12 W. V. O. Quine, *Word and Object, op. cit.*

13 Se presentan los problemas traducidos al español pero las soluciones deben anotarse en inglés. [T.]

14 Como lo expresó el filósofo John Haugeland (*Mind Design,* The MIT Press, Cambridge, 1981, p. 23), el primer principio de la inteligencia artificial es: «Si cuidas la sintaxis, *la semántica se cuidará sola*». Hay diferentes maneras de pensar en este aforismo; en su primera versión, demasiado optimista, motivó la búsqueda de una enorme base de datos, una formalización axiomatizada del conocimiento del mundo que podría mantenerse y explotarse con un motor de inferencia (puramente sintáctica) (cyc es el mejor ejemplo). Esto (a ojos de la mayoría de los expertos) ha demostrado ser inviable, pero el aforismo sobrevive como una buena expresión de la idea de que el cerebro es una clase de computadora (y por lo tanto un motor sintáctico) que, gracias a su diseño, hace aproximadamente el trabajo de un motor semántico.

15 D. Davidson, «Knowing One's Own Mind», *Proceedings and Addresses of the American Philosophical Association,* 60: 441-458, 1987, pp. 443-444.

16 El sistema Watson, de IBM, llegó justo a tiempo para convertir mi ciencia ficción en algo bastante cercano a un hecho científico. Si quieres puedes pensar que Watson está en la caja A, y que la caja B tiene una alternativa sueca que Bo desarrolló de manera independiente. La primera vez que publiqué este experimento mental (D. C. Dennett, *Darwin's Dangerous Idea, op. cit.*) lo mejor que pude decir fue esto: «Para un ejemplo del mundo real de un proyecto así, véase el enorme proyecto cyc (abreviatura de *encyclopedia*) de Douglas Lenat en MCC (Microelectronics and Computer Technology Corporation; D. B. Lenat y R. V. Guha, *Building Large KnowledgeBased Systems, op. cit.*)». Véase también la nota sobre cyc en el apartado 28. Casi nadie había imaginado allá por 1995 el enfoque de la inteligencia artificial que Watson ejemplifica, pero ha estado haciendo grandes progresos. A diferencia de cyc, que está en gran parte codificado «a mano», Watson puede quedarse pastando solo en internet en busca de su propia dotación de verdades y hacer un eficiente uso de las propiedades estadísticas de los datos ahí disponibles. Tanto Watson como cyc, cada uno a su manera, maomeno entienden los datos de sus bancos de datos —muchísimo más que otras computadoras con grandes bases de datos—.

17 La primera vez que escribí esta bomba de intuición, el código ASCII (siglas de *Código Estándar Estadunidense para Intercambio de Información,* en inglés) era el formato estándar para casi todos los textos escritos en procesadores de palabras, el correo electrónico y el lenguaje de internet. Desde entonces ha sido sustituido por un formato extendido, compatible con versiones anteriores: UTF-8 (formato universal de transformación [de conjuntos de caracteres] de 8 bits). Esto es, ASCII aún forma parte de UTF-8.

18 Sólo para filósofos: algunas personas han sostenido que mi explicación de las regularidades en «Real Patterns», *Journal of Philosophy,* 88 (1): 27-51, 1991, es *epifenomenalista* en cuanto al contenido. Esto es mi respuesta.

19 Ya que se les ocurra la hipótesis, los marcianos pueden embarcarse en una versión de la traducción radical de Quine, pero su tarea será doblemente difícil por no poder interrogar a sus «informantes», A y B, poniéndoles objetos enfrente y preguntando «¿pelota?», «¿lápiz?», y así por el estilo.

20 Se puede sacar la misma moraleja de la interpretación de los hechos históricos de la historia evolutiva.

Incluso si puedes describir, en microdetalle sin igual, todos los hechos causales en la historia de cada jirafa que jamás haya vivido, a menos que subas uno o dos niveles y preguntes «¿Por qué?» —en busca de las *razones* que apoyó la Madre Naturaleza—, nunca podrás explicar las regularidades manifiestas —por ejemplo, el hecho de que las jirafas han llegado a tener cuellos largos (nos detendremos más en esto al llegar al capítulo VI)—.

21 Tengo mis propias corazonadas sobre lo que resultará ser una teoría científica del significado apropiada que remplace las rígidas construcciones booleanas (lógicas) de la *inteligencia artificial a la antigua* con redes más flexibles y bayesianas (estadísticas, probabilísticas) de rastreadores de regularidades, pero elaborar esas ideas es una tarea para otro momento y otro lugar.

VI. Herramientas para pensar acerca de la evolución

La idea de Darwin de la evolución por selección natural me parece la mejor idea que nadie haya tenido jamás, porque de un plumazo une el significado con la materia, dos aspectos de la realidad que parecen ser diametralmente opuestos. Por un lado, tenemos el mundo de nuestras mentes y sus significados, nuestros objetivos, nuestras esperanzas y nuestros anhelos, y el tema filosófico más reverenciado —y trillado—, el Sentido de la Vida. Por otro lado, tenemos galaxias girando sin cesar, planetas que se mantienen inútilmente en órbita, mecanismos químicos sin vida haciendo lo que la física ordena, todo sin razón o propósito. Entonces llega Darwin y nos muestra cómo lo primero surge de lo segundo, generando de paso un significado: la visión expansiva de saber por primera vez lo importante que es echar por tierra la desgastada imagen de la tradición. La idea de la selección natural no es muy compleja, pero es tan poderosa que algunas personas no soportan contemplarla y desesperadamente la rehúyen, como si fuera una cucharada de medicina repugnante. Aquí hay algunas herramientas para pensar que nos ayudarán a ver cómo esta idea ilumina los rincones oscuros de la existencia, convierte misterios en acertijos que podemos resolver y descubre las glorias de la naturaleza como nunca antes.

34. El ácido universal

¿Alguna vez has oído hablar del ácido universal? Esta fantasía solía divertirnos a algunos amigos de la escuela y a mí. No tengo idea de si la inventamos o la heredamos, junto con la mosca española y el salitre, como parte de una cultura juvenil subterránea. El ácido universal es un líquido tan corrosivo que perfora y deshace cualquier cosa. ¿Entonces en qué lo guardas? Disuelve botellas de vidrio y botes de acero inoxidable con la misma facilidad que las bolsas de papel. ¿Qué pasaría si de alguna manera crearas o te apoderaras de una porción de ácido universal? ¿Terminaría por destruir todo el planeta? ¿Qué dejaría a su paso? Después de que todo se hubiera transformado tras su encuentro con el ácido universal, ¿cómo se vería el mundo? En ese entonces ni remotamente imaginaba que pocos años después encontraría una idea —la idea de Darwin— que tiene un inconfundible parecido con el ácido universal: corroe prácticamente todos los conceptos tradicionales, y deja a su paso una cosmovisión revolucionada, en la que la mayoría de los viejos puntos de referencia son todavía reconocibles, pero se han transformado de una manera fundamental.[1]

Muchos de los que le tienen pavor a Darwin no entendieron (¿a propósito?) de qué se trataba cuando presenté esta imagen de su peligrosa idea en 1995. Hice todo lo posible para asegurarles a mis lectores que después de que el ácido universal barriera con sus temas favoritos —la ética, el arte, la cultura, la religión, el humor y, en efecto, hasta la conciencia— lo que quedaría sería igual de maravilloso, en muchos aspectos hasta más, pero se habría transformado de manera sutil. La idea de Darwin es revolucionaria, no hay ninguna duda, pero no destruye lo que valoramos en todas estas cosas: les da mejores cimientos y las une elegantemente con el resto del conocimiento. Durante siglos, no sólo se ha considerado que «las artes y las humanidades» están separadas de las ciencias, sino que de algún modo están *protegidas de* los invasivos exámenes de que se ocupa la ciencia, pero este aislamiento tradicional no es la mejor manera de conservar lo que amamos. Tratar de ocultar nuestros tesoros tras un velo de misterio nos impide encontrarles un anclaje adecuado en el mundo físico. Es un error bastante común, sobre todo en filosofía.

Cuando la gente siente que algo que ama está amenazado, su primera reacción es construir un muro «impenetrable», una Línea Maginot, y sólo para estar todavía más seguros decide encerrar un poco más de territorio, una zona colchón, dentro de sus fortificaciones. Parece una idea buena y prudente. Parece protegernos de la horrible pendiente resbaladiza, esa insidiosa *punta de la hebra,* y como todo el mundo sabe, si les das la mano te tomarán el pie. ¡Caven el foso! ¡Levanten el muro! Y tan lejos como nos alcance. Pero este principio normalmente deja a los defensores con la carga de un conjunto de dogmas precario, extravagante (inverosímil, indefendible), que no puede defenderse racionalmente —y por lo tanto al final, en su desesperación, tiene que

defenderse con uñas y dientes—. En filosofía, esta elección estratégica a menudo toma la forma de un *absolutismo* de una u otra clase: la santidad de la vida (humana) es *infinita;* en el corazón del gran arte yace el *genio* divino e inexplicable; la conciencia es un problema demasiado difícil de entender para nosotros los simples mortales, y —uno de mis blancos predilectos— lo que llamo el *realismo histérico:* siempre hay verdades más profundas que resuelven los acertijos del significado. Estas verdades son reales, *realmente* reales, aun cuando seamos sistemáticamente incapaces de descubrirlas. Es una idea tentadora, en parte porque apela a nuestro sentido de apropiada modestia humana. ¿Quiénes somos *nosotros* para decir que simplemente no hay hechos o verdades que resuelvan estos temas? La notoria resistencia de Einstein a la indeterminación de la física cuántica es un ejemplo augusto del atractivo de esta idea. «¡Dios no juega a los dados!», fue su sincero pero en última instancia poco razonable terreno para la resistencia. Y, si te pones a ver, ¿quiénes somos nosotros —o quién es Einstein— para decir que Dios no juega a los dados? Probablemente no sea ésta la manera de abordar estos temas; más adelante veremos más oportunidades para el realismo histérico —y para ver cómo resistirse a él—. El pensamiento evolutivo es un buen antídoto.

35. La biblioteca de Mendel: Extremada y Evanescente

Craig Venter y otros han ya secuenciado el *genoma humano,* pero ¿qué significa eso? ¿Qué no todos tenemos un ADN diferente? Sí, tan diferente, de hecho, que incluso un pequeño fragmento de ADN hallado en la escena de un crimen es suficiente para identificar a su propietario con una certeza de más de 99 por ciento. Y, sin embargo, el ADN humano es también tan similar que los científicos pueden distinguirlo del ADN de otras especies con tan sólo fragmentos de un genoma completo. ¿Cómo es posible? ¿Cómo puede nuestro ADN personal ser tan diferente y a la vez tan similar? Una buena manera de entender este hecho sorprendente es comparando el ADN con los textos de los libros; el escritor argentino Jorge Luis Borges nos ha dado una pequeña fábula, «La biblioteca de Babel», que ilustra vívidamente cómo pueden coexistir esta diferencia y esta similitud.[2] Borges habla de las desesperadas peregrinaciones y especulaciones de la gente que vive en un inmenso almacén de libros, estructurado como un panal, compuesto de miles (o millones o miles de millones) de galerías hexagonales cubiertas de anaqueles, y con pozos de ventilación cercados por barandas. Parados detrás de una baranda viendo hacia arriba o hacia abajo, nadie puede verles un final a estas galerías, y nadie jamás ha encontrado una sola que no esté rodeada de otras seis galerías vecinas. La gente se pregunta si el almacén es infinito. Al final deciden que no lo es, pero bien podría serlo, pues parece que en sus anaqueles —sin ningún orden, ¡ay!— yacen *todos los libros posibles.*

Supón que cada libro tiene 500 páginas, y cada página consiste en 40 renglones, cada uno con 50 espacios, así que hay 2 000 espacios para caracteres por página. Cada espacio o bien es blanco o bien tiene un carácter impreso, elegido de un conjunto de 100 caracteres (las mayúsculas y minúsculas del inglés, el español y otras lenguas europeas, además del espacio y los signos de puntuación).[3]

En algún lugar de la biblioteca de Babel hay un volumen que consiste en puras hojas en blanco, y otro volumen son puros signos de interrogación, pero la inmensa mayoría consisten en incoherencias tipográficas; ninguna regla de ortografía o gramática, por no hablar del sentido, prohíbe que se incluya algún volumen. Quinientas páginas multiplicadas por 2 000 caracteres por página dan un millón de espacios para caracteres por libro, así que si fuéramos a llenar los libros con todas las permutaciones de los caracteres, habría $100^{1\,000\,000}$ libros diferentes en la biblioteca de Babel. Como se calcula que solamente hay 100^{40} (unas más, unas menos) *partículas* (protones, neutrones y electrones) en la región del universo que podemos observar, la biblioteca de Babel no es ni remotamente un objeto físicamente posible, pero gracias a las estrictas reglas con las que Borges la construyó en su imaginación, podemos pensar claramente en ella.[4]

¿Es éste verdaderamente el conjunto de *todos* los libros posibles? Obviamente no,

pues tienen la restricción de que «solamente» usen cien caracteres diferentes, lo que excluye, podemos suponer, los del griego, el ruso, el árabe y el chino, así que pasa por alto muchos de los más importantes libros *reales*. Desde luego, la biblioteca sí contiene magníficas traducciones de estos libros al español, al francés, al alemán, al italiano, etc., así como incontables billones de malas traducciones de cada libro. Hay allí libros de más de 500 páginas que empiezan en un volumen y continúan sin interrupción en otros.

Es divertido pensar en algunos de los volúmenes que deben de estar en alguna parte de la biblioteca de Babel. Uno de ellos es tu mejor biografía de 500 páginas, la más fiel, desde el momento de tu nacimiento hasta tu muerte. Localizarla, sin embargo, sería casi imposible, pues la biblioteca también contiene archimillones de volúmenes que son biografías tuyas magníficamente fieles hasta tu cumpleaños número 10, 20, 30, 40, etc., pero completamente falsas —en un archimillón de maneras diferentes y amenas— en lo que respecta a los sucesos subsiguientes de tu vida. Sin embargo, encontrar aunque sea un volumen entretenido en este inmenso almacén es extremadamente poco probable.

Necesitamos algunos términos para las cantidades implicadas. La biblioteca de Babel no es infinita, así que la probabilidad de encontrar en ella algo interesante no es literalmente infinitesimal.[5] Estas palabras exageran de una manera que nos es conocida, pero deberíamos evitarlas. Desafortunadamente, todas las metáforas habituales —astronómicamente grande, una aguja en un pajar, una gota en el océano— se quedan cómicamente cortas. Ninguna cantidad astronómica *real* (como el número de partículas elementales en el universo o la cantidad de tiempo transcurrido desde el Big Bang medido en nanosegundos) es siquiera visible contra el telón de fondo de estos números enormes pero finitos. Si un volumen ameno en la biblioteca es tan fácil de encontrar como una gota particular en el océano, ¡ya la hicimos! Si nos soltaran al azar en la biblioteca, nuestra oportunidad de encontrar un volumen con tan siquiera una oración gramatical en él es tan evanescentemente pequeña que haríamos bien en escribir el término con mayúscula, Evanescentemente pequeño, y darle una pareja, Extremadamente, abreviatura de Enormemente-más-que-astronómico.[6]

Aquí hay otra manera de darnos una idea de cuán ridículamente grande es la biblioteca de Babel. Como acabamos de observar, sólo un subconjunto Evanescentemente pequeño de los libros están compuestos de palabras en inglés. Ese subconjunto es de suyo Extremo, y un Evanescente subconjunto dentro de él consiste en libros en los que las palabras están ordenadas en oraciones gramaticales (la Extrema mayoría consiste en volúmenes llenos con cadenas de palabras como ésta: «bueno desde París ayuda fácil de en la cual obstante no democracia encuera tigres»). Un subconjunto Extremo pero Evanescente de los libros gramaticales están compuestos de oraciones que tienen sentido en secuencia (el resto están compuestos de oraciones que podrían elegirse al azar de libros en español gramatical). Un subconjunto Extremo pero Evanescente de

esos libros que tienen sentido son acerca de alguien que se llama John, y un subconjunto Extremo pero Evanescente de *ésos* son acerca del asesinato de John F. Kennedy, y un subconjunto todavía Extremo (pero Evanescente) de *ésos* son verdaderos, ¡y un subconjunto Extremo pero Evanescente de esos libros verdaderos acerca del asesinato de Kennedy están escritos con puros epigramas! Sí, ¡hay más libros verdaderos en epigramas acerca de la muerte de J. F. K. *posibles* que volúmenes en la Biblioteca Nacional! Con toda probabilidad, ni uno solo se ha publicado, y menos mal que ha sido así.

Moby Dick está en la biblioteca de Babel, pero también 100 000 000 de impostores mutantes que difieren del *Moby Dick* canónico por un *solo* error tipográfico. Eso no llega a ser un número Extremo, pero el total aumenta rápidamente cuando añadimos las variaciones que difieren por dos, 10 o 1 000 erratas. Hasta un volumen con 1 000 erratas —dos por página en promedio— sería inconfundiblemente *Moby Dick,* y de

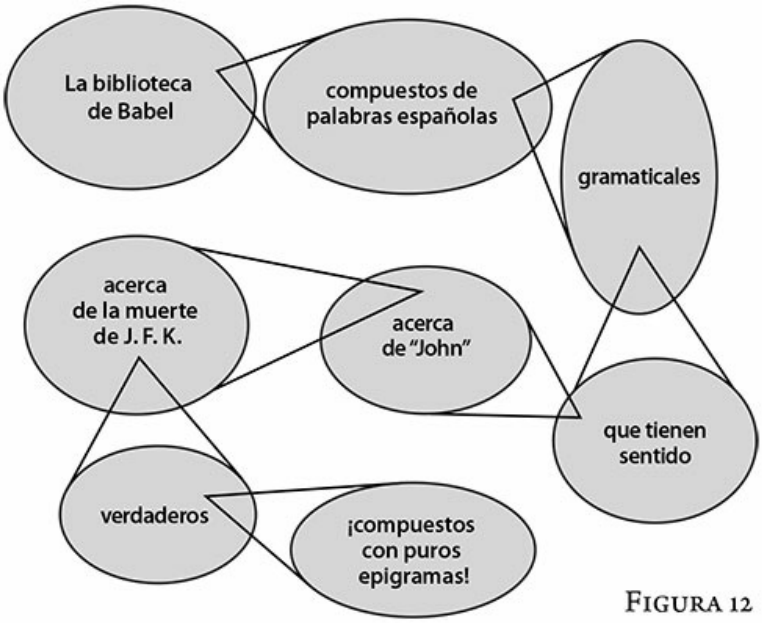


FIGURA 12

esos volúmenes hay un número Extremadamente grande. No importaría cuál de ellos encontraras, ¡si tan sólo pudieras encontrar uno! Casi todos serían una lectura igualmente maravillosa, y cuentan la misma historia, salvo por unas diferencias verdaderamente insignificantes, casi indiscernibles. No todas, sin embargo. A veces una sola errata, en una posición decisiva, puede ser fatídica. Peter De Vries, otro escritor deliciosamente filosófico, publicó una novela[7] que empezaba: «Llámeme, Ismael». ¡Oh, lo que puede hacer una sola coma! O piensa en los muchos mutantes que empiezan con «Lámeme Ismael».

En el cuento de Borges los libros no están guardados en los anaqueles con ningún orden, pero aunque estuvieran en un escrupuloso orden alfabético, tendríamos problemas

irresolubles para encontrar *el* libro que estuviéramos buscando (por ejemplo, la versión «esencial» de *Moby Dick*). Imagínate viajar en un cohete espacial a través de la galaxia Moby Dick de la biblioteca de Babel. Esta galaxia es de por sí Extremadamente más grande que todo el universo físico, así que no importa en qué dirección viajes, siglos y siglos, aunque viajes a la velocidad de la luz, todo lo que ves son ejemplares prácticamente indistinguibles de *Moby Dick*. Nunca alcanzarás nada que parezca algo distinto. *David Copperfield* está inimaginablemente lejos en este espacio, aunque sabemos que hay un camino —el camino más corto, ignorando los otros archimillones— que lleva de un libro al otro mediante un solo cambio tipográfico (si te encontraras en este camino, te resultaría casi imposible saber, revisando el lugar, qué dirección tomar hacia *David Copperfield,* aunque tuvieras textos de los dos libros buscados a la mano).

En otras palabras, este espacio *lógico* es tan Extremadamente grande que muchas de nuestras ideas habituales acerca de la ubicación, del examinar, del encontrar y otras actividades mundanas y prácticas por el estilo, no tienen una aplicación sencilla. Borges puso los libros en los anaqueles en orden aleatorio, un toque simpático del que derivó varias reflexiones exquisitas, pero mira los problemas que él mismo se habría buscado si hubiera tratado de arreglarlos en orden alfabético. Como sólo hay cien caracteres alfabéticos diferentes (en nuestra versión), podemos tratar alguna secuencia específica de ellos en orden alfabético, por ejemplo: a, A, b, B, c, C, … z, Z, ?, ;, ,, ., !,), (, %, … à, â, á, è, ê, é, … Entonces podemos poner todos los libros que empiezan con el mismo carácter en el mismo *piso*. Ahora nuestra biblioteca sólo tiene 100 pisos: ni siquiera es tan alta como la Torre Sears (o Willis) en Chicago. Podemos dividir cada piso en 100 corredores, y cubrimos cada uno con los libros cuyo segundo carácter es el siguiente carácter en orden alfabético. En cada corredor podemos colocar 100 anaqueles, uno por cada lugar del tercer carácter. Así, todos los libros que empiezan con «Aarónicos amantes de Mozart» —¡y hay tantos!— están colocados en el mismo anaquel (el anaquel de las «r») del primer corredor del primer piso. Pero ése es un anaquel inmensamente largo, así que tal vez mejor debamos amontonar los libros en archiveros acomodados en ángulo recto con respecto al anaquel, un cajón para cada posición de cuarta letra. De esa manera, cada anaquel puede medir sólo, digamos, 30 metros. Pero ahora los archiveros tienen demasiado fondo y van a chocar con la parte posterior de los archiveros del corredor vecino, así que… Pero ya nos quedamos sin dimensiones donde formar los libros. Necesitamos un espacio de millones de dimensiones para guardarlos todos ordenadamente, y lo único que tenemos son tres dimensiones: arriba-abajo, izquierda-derecha y frente-reverso. Así que simplemente tendremos que fingir que podemos imaginar un espacio multidimensional, y que cada dimensión se cruce en «ángulos rectos» con todas las demás. Podemos concebir tales hiperespacios, como se les llama, aunque no podamos visualizarlos. Los científicos los usan todo el tiempo para organizar

la expresión de sus teorías. La geometría de esos espacios (cuenten o no cuenten como sólo imaginarios) se sabe comportar y los matemáticos la han explorado mucho. Podemos hablar con confianza de ubicaciones, caminos, trayectorias, volúmenes (hipervolúmenes), distancias y direcciones en estos espacios lógicos.

Ya estamos preparados para considerar una variación sobre el tema de Borges, que llamaré *la biblioteca de Mendel.* Esta biblioteca contiene «todos los posibles genomas» —secuencias de ADN—. Richard Dawkins describe un espacio similar, que llama Tierra Biomórfica, en *El relojero ciego.*[8] Su discusión inspiró la mía, y nuestras dos explicaciones son completamente compatibles, pero quiero hacer hincapié en algunos aspectos que él decidió pasar por alto.

Si consideramos que la biblioteca de Mendel está compuesta de *descripciones* de genomas, entonces ya es una parte adecuada de la biblioteca de Babel. El código estándar para describir ADN consta de sólo cuatro caracteres, A, C, G y T (por adenina, citosina, guanina y timina, las cuatro clases de nucleótidos). Por lo tanto, todas las permutaciones de 500 páginas de estas cuatro letras ya están en la biblioteca de Babel. Sin embargo, los genomas típicos son mucho más largos que los libros comunes y corrientes. El genoma humano tiene aproximadamente 3 000 millones de nucleótidos, así que la descripción exhaustiva de un solo genoma humano —como el tuyo propio— ocuparía aproximadamente 3 000 volúmenes de 500 páginas de la biblioteca de Babel.

Esta comparación de un genoma humano con los volúmenes de la galaxia de *Moby Dick* ahora nos da la explicación que necesitamos de las diferencias y similitudes entre genomas humanos. ¿Cómo podemos hablar de secuenciar (copiar) *el* genoma humano si todo genoma humano es diferente de todos los demás no sólo en un lugar (*locus,* en el lenguaje de los genetistas), sino en cientos o miles? Como los proverbiales copos de nieve, o las huellas digitales, no hay dos genomas humanos que sean exactamente iguales, ni siquiera los de gemelos idénticos (siempre se corre el riesgo de que se deslice alguna errata, incluso en las células de un solo individuo). El ADN humano se distingue fácilmente del de cualquier otra especie, hasta del ADN del chimpancé, que es igual en más de 90 por ciento de los *loci*. Una galaxia de genomas humanos posibles que está Extremadamente lejos de las galaxias de genomas de otras especies contiene todos los genomas humanos que realmente han existido. Sin embargo, dentro de la galaxia hay espacio suficiente para que no haya dos genomas humanos parecidos. Tienes dos versiones de cada uno de tus genes, uno de tu madre y otro de tu padre. Ellos te pasaron exactamente la mitad de sus propios genes, aleatoriamente seleccionados de entre los que ellos recibieron de sus padres, tus abuelos, pero como tus abuelos también eran miembros de la especie *Homo sapiens,* sus genomas concuerdan en casi todos los *loci,* así que la mayor parte del tiempo da lo mismo qué abuelo te proporcione cualquiera de tus genes. No obstante, sus genomas difieren en muchos miles de *loci,* y en esos

espacios, qué genomas obtienes es cuestión de azar: un volado incorporado a la maquinaria para formar las contribuciones de tus padres a tu ADN. Por otra parte, las mutaciones en los mamíferos se acumulan aproximadamente a razón de 100 por genoma por generación. «Esto es, tus hijos tendrán 100 diferencias en sus genes con respecto a ti y a tu cónyuge como resultado de errores aleatorios de copiado por tus enzimas o como resultado de mutaciones en tus ovarios o testículos causadas por rayos cósmicos.»[9]

La descripción del genoma de un caballo, una col o un pulpo estaría compuesta de las mismas letras, A, C, G y T. La mayoría de los genomas animales que se han medido son más pequeños que el humano, pero algunas plantas tienen genomas que son más de 10 veces mayores que el nuestro, ¡y algunas amibas unicelulares tienen genomas todavía más grandes! *Amoeba dubia* tiene el récord mundial actual, con un cálculo aproximado de 670 mil millones de pares de bases, más de 200 veces más grande que el nuestro. Pero supongamos, arbitrariamente, que la biblioteca de Mendel consta de todas las cadenas de ADN descritas en todos los estuches de 3 000 volúmenes que constan únicamente de esos cuatro caracteres. Los «posibles» genomas que así se capten serán suficientes para cualquier propósito teórico serio.

Exageré al decir que la biblioteca de Mendel contiene «todos los genomas posibles». Tal como la biblioteca de Babel ignoraba los idiomas ruso y chino, la biblioteca de Mendel ignora la (aparente) posibilidad de alfabetos genéticos alternativos, basados por ejemplo en diferentes constituyentes químicos. Así, puede ser que cualquier conclusión que saquemos con respecto a lo que es posible en relación con *esta* biblioteca de Mendel tenga que reconsiderarse cuando tratemos de aplicarla a alguna noción de posibilidad más amplia. Esto es de hecho una fortaleza más que una debilidad de nuestra táctica, pues podemos vigilar exactamente de qué clase de posibilidad modesta y limitada estamos hablando.

Una de las características importantes del ADN es que todas las permutaciones de secuencias de adenina, citosina, guanina y timina son más o menos igual de estables desde el punto de vista químico. Todas podrían estar elaboradas, en principio, en el laboratorio de acoplamiento de genes, y una vez elaboradas tendrían una vida de anaquel indefinida, como un libro en una biblioteca. Pero no todas estas secuencias de la biblioteca de Mendel corresponden a un organismo viable. La mayoría de las secuencias de ADN —la Extrema mayoría— son incoherencias: recetas para absolutamente nada viviente. Todos los genomas que vemos, que realmente existen hoy en día, son producto de miles de millones de años de ajuste y revisión, un ciego proceso editorial que es efectivo porque la mayor parte de las incoherencias (todas salvo una hebra Evanescentemente delgada de «texto» útil y significativo) se descarta automáticamente, mientras que el resto incesantemente se usa y se copia archimillones de veces. Ahora mismo tienes más de un billón de copias de tu genoma en tu propio cuerpo, una copia en

cada célula humana, y todos los días, mientras se hacen nuevas células de piel, de huesos y de sangre, se instalan en ellas nuevas copias de tus genomas. El texto que *puede* copiarse —por residir en un asunto en marcha: una célula viva— se copia. El resto se disuelve. Publicar o perecer.

36. Los genes como palabras o como subrutinas

La analogía entre los genes y las palabras es útil, como acabamos de ver, pero hay una todavía mejor que ahora estamos preparados para entender gracias al intermedio sobre computadoras. En su obra maestra, *El cuento del antepasado,* Richard Dawkins da a otro inteligente autor que ha escrito sobre la evolución, Matt Ridley, el crédito por haber señalado en su libro *Nature via Nurture* las hondas similitudes entre los genes y subrutinas de *software.* Normalmente no incluiría en un libro mío una cita tan larga de la obra de alguien más, pero me di cuenta de que mis intentos de parafrasear este pasaje invariablemente sacrificaban claridad y viveza por una pizca de originalidad, así que, con su permiso, aquí les paso la versión sin mutaciones tal como aparece en el libro de Dawkins.

> La mayor parte del genoma que secuenciamos no es el libro de instrucciones, o el programa de computadora maestro, para construir un humano o un ratón (aunque unas porciones sí lo son). Si lo fuera, en efecto esperaríamos que nuestro programa fuera más grande que el del ratón. Sin embargo, la mayoría del genoma se parece más al diccionario de palabras disponibles para escribir el libro de instrucciones —o, como pronto veremos, al conjunto de subrutinas a las que recurre el programa maestro—. Como dice Ridley, la lista de palabras en *David Copperfield* es casi la misma que la de *El guardián en el centeno.* Ambos se sirven del vocabulario de un hablante nativo del inglés que ha ido a la escuela. La diferencia entre los dos libros es el orden en el que esas palabras se juntan unas con otras.
>
> Cuando se hace una persona, o cuando se hace un ratón, las dos embriologías se sirven del mismo diccionario de genes: el vocabulario normal de las embriologías de los mamíferos. La diferencia entre una persona y un ratón proviene del distinto orden en que los genes, provenientes de ese vocabulario mamífero compartido, se emplean, los diferentes lugares del cuerpo donde esto ocurre y su sincronización. Todo eso está a cargo de genes particulares cuyo trabajo consiste en encender otros genes, en cascadas complicadas y exquisitamente sincronizadas. No obstante, esos genes directores constituyen sólo una minoría de los genes que tiene el genoma.
>
> No se malinterprete «orden» como si significara el orden en el que los genes están formados a lo largo de los cromosomas. Con notables excepciones […] el orden de los genes en un cromosoma es tan arbitrario como el orden en el que se listan las palabras en un vocabulario. Normalmente es alfabético, pero, sobre todo en libros de frases para viajar a otros países, a veces tienen un orden que responde a la conveniencia: palabras útiles en aeropuertos, palabras útiles cuando vas al médico, palabras útiles para ir de compras, etc. El orden en que los genes se guardan en los cromosomas no tiene importancia: lo que importa es que la maquinaria celular

encuentre el gen correcto cuando lo necesita, y lo hace con métodos que cada vez entendemos mejor […]

La analogía con las palabras es equívoca en un aspecto. Las palabras son más chicas que los genes, por lo que algunos escritores han comparado cada gen con una oración. Sin embargo, las oraciones no son una buena analogía, por distinta razón: no se juntan dos libros permutando un repertorio fijo de oraciones. La mayoría de las oraciones son singulares. Los genes, como las palabras pero a diferencia de las oraciones, se usan una y otra vez en distintos contextos. Una mejor analogía para un gen que una palabra o una oración es una subrutina de caja de herramientas en una computadora […]

La Mac tiene una caja de herramientas de rutinas almacenada en la memoria ROM (*read only memory,* memoria de sólo lectura) o en archivos del sistema cargados perma nentemente en el periodo de arranque. Hay miles de estas rutinas de caja de herramientas, y cada una hace una operación particular, que es probable que se necesite una y otra vez de maneras ligeramente distintas, en diferentes programas. Por ejemplo, la rutina de caja de herramientas llamada ObscureCursor esconde el cursor de la pantalla hasta la siguiente vez que mueves el ratón. Sin que tú lo veas, se convoca al gen de ObscureCursor cada vez que empiezas a teclear y el cursor del ratón se desvanece. Hay rutinas de caja de herramientas detrás de las características familiares que comparten todos los programas de la Mac (y sus imitaciones equivalentes en las máquinas que usan Windows): menús desplegables, barras de desplazamiento, ventanas reducibles que puedes arrastrar por la pantalla y muchas otras.[10]

Todo esto nos ayuda a entender por qué es tan fácil para un experto reconocer el genoma de un mamífero. Éste tiene la caja de herramientas de los mamíferos, que además de sus herramientas especializadas en hacer mamíferos incluye herramientas de las cajas de los reptiles, de los peces y hasta de los gusanos. Todos los seres vivos, entre ellos las bacterias, tienen en común las herramientas más antiguas del estuche.

37. El árbol de la vida

Los genomas que existen hoy en día están conectados por hebras de ascendencia con los genomas de sus padres, abuelos, etc., hasta el principio de la vida en la Tierra. La figura 13, el árbol de la vida, muestra cómo *cada* persona está relacionada de manera relativamente cercana a toda otra persona: tienen en común ancestros humanos que vivieron en los últimos 100 000 años, y también los tienen con todos los perros y ballenas que vivieron en los últimos 200 millones de años, y con toda margarita y toda secuoya que haya vivido durante los últimos 2 000 millones de años.

Hay muchas maneras de dibujar el árbol de la vida. En ésta, el presente se representa a lo largo de las puntas externas de las ramas. Sólo los linajes que siguen vivos alcanzan este aro externo. Se muestra que los dinosaurios (salvo por las aves, que descendieron de ellos) se extinguieron hace más de 60 millones de años. Todas las hebras se conectan en algún momento y se remontan al origen de la vida (en lo que abundaremos más adelante). Si agrandáramos el diagrama un billón de veces, veríamos los árboles genealógicos de todas las moscas, peces y ranas que jamás hayan vivido, y veríamos cuáles murieron sin descendencia (la mayoría, desde luego) y cuáles tuvieron progenie.

38. Grúas y ganchos celestes: levantamiento en el Espacio de Diseño

La vida es asombrosa. Cuando piensas en los miles de millones de sistemas solares que casi con seguridad carecen por completo de vida, es asombroso que simplemente haya alguna manera de estar vivo, y es asombroso reflexionar sobre la variedad de maneras de estar vivo, de las bacterias a los peces, pájaros, margaritas, ballenas, víboras, arces y personas. A lo mejor lo más asombroso de todo es la tenacidad de las cosas vivientes, los miles de maneras que tienen de aferrarse a la vida y reproducirse, ganándose la vida a duras penas y en contra de obstáculos formidables, gracias a millones de recursos y arreglos ingeniosos, desde las cascadas enrolladas de maquinaria proteínica dentro de cada célula, a la ecolocalización de los murciélagos, la trompa del elefante, la capacidad de nuestros cerebros para reflexionar sobre todos los temas «bajo el sol» y muchos otros también. Todo ese magnífico ajuste de los medios a los fines requiere una explicación, pues no puede ser puro azar o casualidad. Sólo hay dos posibilidades conocidas: diseño inteligente o evolución por la selección natural. En cualquiera de los dos casos hay una tremenda cantidad de trabajo de diseño por hacer, ya sea milagrosamente a manos de un Diseñador Inteligente, o lenta, pesada, imprevisible y estúpidamente —pero no milagrosamente— por la selección natural. Llamemos al trabajo de diseño *I+D* (por investigación y desarrollo) y reconozcamos que eso siempre es costoso. Toma tiempo y energía. El encanto de la gran idea de Darwin es que vio cómo, dados miles de millones de años para trabajar, y una prodigiosa cantidad de movimiento «desperdiciado» (archimillones de ensayos que terminan en error), las mejoras en el diseño pudieron acumularse automáticamente, de manera no milagrosa, sin intención, previsión ni entendimiento. Robert Beverley MacKenzie, uno de los críticos más fervientes de Darwin, lo expresa de manera elocuente:

> En la teoría con la que tenemos que lidiar, la Ignorancia Absoluta es el artífice, así que podemos enunciar como el principio fundamental de todo el sistema que PARA FABRICAR UNA MÁQUINA PERFECTA Y BELLA NO HACE FALTA SABER CÓMO HACERLA. Si se estudia con cuidado, se verá que esta proposición expresa, de manera condensada, el propósito esencial de la Teoría, y expresa en pocas palabras todo el sentido del señor Darwin, quien, por una extraña inversión del razonamiento, parece pensar que la Ignorancia Absoluta está plenamente calificada para tomar el lugar de la Sabiduría Absoluta en todos los logros de la habilidad creativa.[11]

Exactamente. Y desde que Darwin propuso esta idea asombrosa, los escépticos se han preguntado si hubo tiempo suficiente para hacer todo ese trabajo creativo de la

manera lenta y pesada. Para imaginar el trabajo de diseño que necesita hacerse puede pensarse en él como *levantamiento* en el Espacio de Diseño. ¿Qué es el Espacio de Diseño? Como la biblioteca de Babel y la de Mendel, la mejor manera de concebirlo es como un espacio multidimensional. De hecho, el Espacio de Diseño contiene las dos bibliotecas y más, porque no sólo incluye todos los libros (diseñados, escritos por un autor) y los organismos (diseñados, evolucionados), sino todas las otras cosas que pueden describirse bien desde la postura de diseño (véase el apartado 18), como casas, ratoneras, hachas de guerra, computadoras y cohetes espaciales. Así como la mayor parte de la biblioteca de Babel son incoherencias, la mayoría de los lugares en el Espacio de Diseño están llenos de basura, cosas que no pueden hacer nada bien. Si eres como yo, puedes imaginar sólo tres dimensiones simultáneamente, pero mientras más juegas con la idea en tu imaginación más fácil se vuelve pensar en las tres dimensiones conocidas como si sustituyeran a muchas (esta herramienta de pensamiento mejora con la práctica).

Cuando imaginamos colocar el árbol de la vida en la biblioteca de Mendel podemos ver cómo todas las cosas vivientes que jamás han existido realmente en este planeta están conectadas entre sí con líneas de descendencia. Esos linajes transmiten las mejoras de diseño básicas que la selección natural ha «descubierto», y conserva la más antigua I+D para que la usen todas las subsiguientes cosas vivientes. (La «maquinaria» dentro de cada célula, desde las bacterias hasta las neuronas, incluye miles de nanodispositivos de brillante diseño que han estado dándole vuelta a la manivela durante más de 3 000 millones de años: el cuarto de máquinas básico de la vida, que tenemos en común con los árboles, los pájaros y las células de levadura.) En el nivel multicelular encontramos corazones, pulmones, ojos, brazos, piernas y alas, para los que se han reutilizado y mejorado diseños que se desarrollaron en primer lugar hace «apenas» 2 000 millones de años. Además de partes de organismos, están los artefactos de los organismos: telarañas, nidos y los diques que construyen los castores, por ejemplo. Éstos también dan muestra inconfundible de I+D, y sus frutos tienen que haberse transmitido de alguna manera a lo largo de las mismas líneas de ascendencia o descendencia familiar: ya sea en los genes o con hijos que imitan a sus padres. Cada cierto tiempo surge una novedad —por mutación, experimentación o accidente— que representa una mejora, y se copia, copia y copia. Los experimentos fallidos se extinguen. Otra vez: publicar o perecer.

Y luego tenemos los artefactos humanos, arados, puentes y catedrales, pinturas, obras de teatro, poemas, y las prensas de imprenta, los aviones, las computadoras y las podadoras, y… las herramientas de pensamiento. ¿Qué decir de ellas? ¿No cuelgan de alguna de las ramas de una parte del árbol de la vida: el linaje humano? Cada una depende al menos de un autor o inventor, y la mayoría de ellas dependen de incontables miles o millones de colaboradores de la I+D que tienen detrás. Beethoven no tuvo que inventar la sinfonía, que ya estaba ahí para que él la adaptara, y Shakespeare no tuvo

que inventar el soneto. Una sierra eléctrica está hecha de decenas o cientos de elementos «ya hechos», que ya se inventaron y optimizaron. Algunos de nuestros artefactos humanos probablemente se copiaron de los artefactos de otros animales. ¿El nido del ave tejedora inspiró la invención de las telas tejidas? La plataforma elevada que evita que se filtre el frío por la puerta en el iglú de un esquimal, ¿se copió del piso elevado similar de las guaridas de los osos polares, o son invenciones independientes (o los osos polares lo copiaron de los esquimales)?

Las profundas semejanzas entre las funciones de los corazones y las bombas, la ecolocalización de los murciélagos y el sonar y el radar, los diques de los castores y los de irrigación, los ojos y las cámaras, no es accidental: procesos de exploración — de I+D — parecidos les han dado forma y los han pulido y mejorado a lo largo de los siglos. Ponlos todos en un espacio común: el Espacio de Diseño, ese espacio de *todos los diseños posibles*.

Esta movida es polémica entre los biólogos, por razones que creo entender y que lamento. Muchos biólogos son enormemente reacios a hablar de «diseño» en las cosas vivientes, porque piensan que eso apoya y reconforta al movimiento del diseño inteligente, esa campaña falsa, seudocientífica, criptorreligiosa para minar la ampliamente merecida autoridad de la biología evolutiva. (Entre estos reacios están dos de mis colegas y amigos más queridos, los biólogos evolutivos Richard Dawkins y Jerry Coyne.) Al hablar de propósito y diseño en la naturaleza (aparentemente) le concedemos a la banda del diseño inteligente la mitad de sus argumentos; algunos piensan que es mejor mantener una rígida prohibición sobre esos temas e insistir en que *estrictamente hablando* nada en la biosfera está diseñado a menos que artífices humanos lo hayan diseñado. La manera que tiene la naturaleza para generar sistemas complejos (órganos, comportamientos) es tan diferente a la de un artífice, piensan, que no deberíamos usar el mismo lenguaje para describirlos. Así, Richard Dawkins de vez en cuando habla de características *diseñoides* en los organismos,[12] y en *El cuento del antepasado* dice: «La ilusión del diseño conjurada por la selección natural darwiniana es de una potencia pasmosa».[13] Estoy en desacuerdo con este principio de austeridad, que puede fracasar seriamente. Hace poco, en un bar, no pude evitar oír una conversación entre unos jóvenes acerca de las maravillas de la nanomaquinaria que se ha descubierto dentro de todas las células. «Cuando ves a todos esos fantásticos pequeños robots trabajando, ¡cómo puedes creer en la evolución!», exclamó uno, y otro asintió con la cabeza con aire de entendido. De algún modo, estas personas tenían la impresión de que los biólogos evolutivos piensan que la vida no es tan compleja, no está hecha de componentes tan maravillosos. Estos escépticos de la evolución no eran pueblerinos reaccionarios: ¡eran estudiantes de medicina de Harvard! Subestimaban enormemente el poder de la selección natural porque los biólogos evolutivos les habían dicho, una y otra vez, que no hay diseño *real*

en la naturaleza, sólo la *apariencia* de diseño. El episodio me convenció de que los «conocimientos del dominio público» empiezan a incorporar la idea equivocada de que los biólogos evolutivos son reacios a «admitir» o «reconocer» todo el diseño evidente que hay en la naturaleza.

Considera a este respecto a Christoph Schönborn, el arzobispo católico de Viena, el tipo al que embaucaron los partidarios del diseño inteligente. Es conocido por lo que escribió en un artículo de opinión en el *New York Times* titulado «Finding Design in Nature» [«Encontrar el diseño en la naturaleza»]:

> La Iglesia católica, si bien le deja a la ciencia muchos detalles sobre la historia de la vida en la Tierra, proclama que el intelecto humano, mediante la luz de la razón, puede clara e inmediatamente discernir propósito y diseño en el mundo natural, que incluye el mundo de las cosas vivientes. Puede ser que la evolución, en el sentido de que hay ancestros comunes, sea verdadera, pero la evolución en el sentido neodarwiniano —un proceso no guiado, no planeado, de variación aleatoria y selección natural— no lo es. Cualquier sistema de pensamiento que niegue o trate de refutar las abrumadoras pruebas de que hay diseño en la biología es ideología, no ciencia.[14]

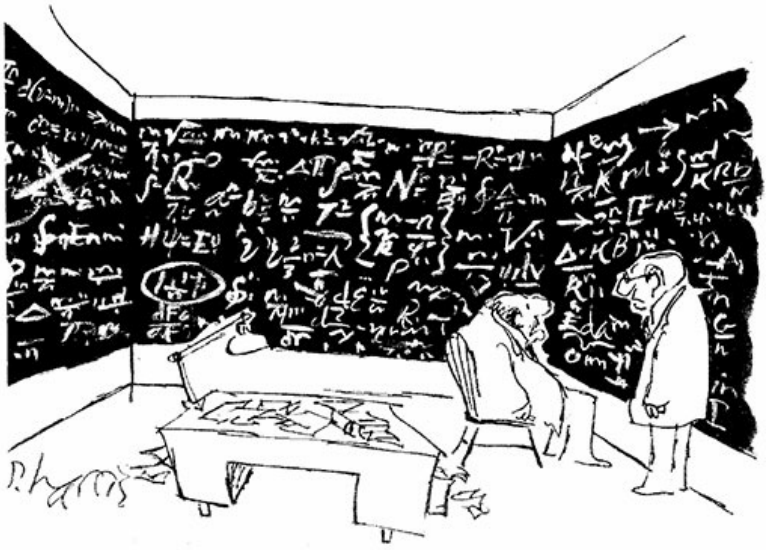
¿Qué campaña queremos liderar los evolucionistas? ¿Queremos tratar de convencer a los no expertos de que en realidad no ven el diseño que es apabullantemente obvio en la biología en todos los niveles, o más bien preferimos mostrar que —es maravilloso decirlo — lo que Darwin ha mostrado es que puede haber diseño —diseño real, real a más no poder— sin un Diseñador Inteligente? Hemos convencido al mundo de que la Tierra gira alrededor del Sol y de que el tiempo es relativo, no absoluto. ¿Por qué rehuir la tarea pedagógica de mostrar que puede haber diseño sin diseñador? Así, pues, aquí estoy defendiendo (una vez más, con un nuevo acento) la siguiente afirmación:

La biosfera está completamente saturada de diseño, de propósito, de razones. Lo que llamo la postura del diseño predice y explica características presentes a lo largo del mundo viviente mediante las mismas suposiciones que funcionan tan bien cuando aplicamos ingeniería inversa a artefactos hechos por diseñadores humanos (un poco) inteligentes. La evolución por la selección natural es un conjunto de procesos que «encuentran» y «rastrean» razones para que las cosas se configuren de una manera más que de otra. La principal diferencia entre las razones que encuentra la evolución y las razones que encuentran los diseñadores humanos es que estas últimas normalmente (pero no siempre) se representan en las mentes de los diseñadores, mientras que las razones reveladas por la selección natural normalmente los primeros que las representan son los investigadores humanos que consiguen hacerles ingeniería inversa a las producciones de

la naturaleza. Es decir, los diseñadores humanos *piensan acerca* de las razones de las características de sus artefactos, y por lo tanto *tienen ideas* que representan las razones. Normalmente observan, aprecian, formulan, refinan, y luego transmiten, discuten, critican las razones de sus diseños. La evolución no hace nada de esto; sólo pasa ciegamente por el tamiz la variación que genera, y lo bueno (que es bueno en virtud de razones, razones con las que el proceso de selección natural no sueña siquiera y tampoco se las representa) se copia.

Podría haber un buen uso del término *diseñoide* para señalar lo que es simple diseño aparente, pero no en biología. Cuando pienso en la simple apariencia de diseño, pienso en cómo los caricaturistas montan el escenario para sus chistes acerca de científicos mostrando a unos sabihondos barbones parados frente a un pizarrón cubierto de símbolos (que de hecho no tienen sentido), en un laboratorio de química con una selva de tubos y matraces o mostrándonos al inventor loco trabajando en su máquina del tiempo, erizada de cuadrantes, antenas y aparatitos de alta tecnología. Estas cosas no podrían hacer nada; sólo parecen ser funcionales. Pero los diseños de la naturaleza son verdaderamente efectivos. De hecho, se puede demostrar que a menudo son mucho más eficientes y poderosos que cualquier contraparte humana que se haya inventado.

Un ejemplo de inscripción diseñoide:



«¿Dónde quedaron las soluciones *elegantes*?»
(Cortesía de Sidney Harris, <sciencecartoonsplus.com>.)

«¿Dónde quedaron las soluciones *elegantes*?» (Cortesía de Sidney Harris, <sciencecartoonsplus.com>.)

Un hombre joven y sano puede caminar, llevando consigo toda la comida y el agua que necesite para sobrevivir, más o menos durante una semana, deteniéndose y descansando cada vez que quiera, y avanzar aproximadamente 250 kilómetros (el agua es la carga determinante: 20 kilos de agua, 7 kilos de comida y 5 kilos de bártulos son un

equipaje muy pesado; si pudiera encontrar agua en el camino, podría seguir meses enteros). Para comparar, considera la creación del especialista en robótica de la Cornell University Andy Ruina y sus colegas, que han construido el que por mucho es el robot que más ha caminado, Ranger, que viajó sin parar 65.2 kilómetros en un ultramaratón de robots en Japón el 1 y el 2 de mayo de 2011. Los diseñadores de Ranger aprovechan las propiedades dinámicas de las extremidades para crear un caminador con una magnífica eficiencia energética (que da vueltas y vueltas horas enteras en una pista plana, dirigido por un ser humano con una palanca de mando). Otro magnífico robot caminante de cuatro patas, Big Dog, es como 15 veces menos energéticamente eficiente, pero tiene capacidades mucho más admirables para adaptarse a un terreno difícil. Los humanos siguen siendo dispositivos de transporte considerablemente más eficientes que Ranger por un factor de cuatro o cinco, y a diferencia de Ranger, pueden responder de manera autónoma a toda clase de características del mundo.[15]

Algo tiene que concederse. O bien *definimos* «diseño» como el producto de un Diseñador Inteligente, como el diseñador de poemas o coches, por ejemplo, o reconocemos que puede haber diseño —genuino diseño— sin un Diseñador Inteligente. La tradición y la etimología parecerían favorecer la primera opción, pero piensa en lo siguiente. Átom*o* (del griego *a* [sin] + *tomos,* de *temnein,* cortar o rebanar) originalmente significaba «cosa indivisible», pero la ciencia descubrió que, a fin de cuentas, dividir el átomo no era una contradicción en sus términos. Sostengo que la ciencia también ha descubierto que el diseño sin diseñador (sin un diseñador con una mente, previsión e intención) no sólo es posible, sino que está en todas partes. El diseño por evolución es un proceso real y bien entendido. Difiere de manera interesante del diseño por ingeniero, pero también es sumamente parecido y puede darle una lección de humildad al «ingenio». Una y otra vez, biólogos desconcertados ante alguna pieza aparentemente inútil o torpe de mal diseño en la naturaleza tarde o temprano han caído en la cuenta de que habían subestimado el auténtico resplandor de una de las creaciones de la Madre Naturaleza. Francis Crick maliciosamente bautizó esta tendencia con el nombre de un colega, Leslie Orgel, al hablar de lo que llamó la Segunda Ley de Orgel: «La evolución es más lista que tú». (Esta táctica aparentemente insensata de personificar el proceso de la selección natural como la Madre Naturaleza, este antropomorfismo descarado, se defenderá a su debido tiempo. No es sólo una metáfora divertida: es una herramienta de pensamiento por derecho propio.)

De vuelta al Espacio de Diseño, el espacio multidimensional de todos los diseños *posibles,* entre ellos no sólo los organismos, automóviles y poemas, sino —como los libros en la biblioteca de Babel— todos los diseños que nunca evolucionaron y nunca se construyeron: los canguros parlantes, las víboras voladoras, las máquinas nucleares para hacer palomitas y los patines submarinos. ¿Cuáles deberían ser los elementos de nuestro

alfabeto de diseño básico? Como no vamos a tratar de *hacer* el Espacio de Diseño, sino sólo pensar acerca de él, podemos ser despilfarradores, ¿por qué no?: digamos, todas las configuraciones posibles de átomos, provenientes de la tabla periódica de los elementos. (Este espacio Extremo incluirá también duplicados átomo por átomo de todas las cosas *no diseñadas,* como todos los guijarros de todas las playas y el monte Everest, pero así es como debe ser, pues nada impide que alguien diseñe y construya réplicas de estas entidades.) ¿En qué parte del Espacio de Diseño estará la *Tercera Sinfonía* de Brahms? Veamos: la *partitura* (con motas de tinta sobre el papel y demás) estará en archimillones de lugares, y las grabaciones en casete y discos de vinil estarán en otros, así que con toda seguridad estará en el Espacio de Diseño. Una canción que nunca se haya escrito o grabado pero que de algún modo se conserve y se haya transmitido por tradición oral será más difícil de atar a familias de átomos, pero con el tiempo una subsección convenientemente intrincada del Espacio de Diseño la alojará bastante bien. Así como la biblioteca de Babel está llena sobre todo de basura, de libros difíciles de leer llenos de disparates, el Espacio de Diseño está lleno principalmente de cosas disparatadas carentes de interés, sin absolutamente ninguna función o capacidad, pero aquí y allá relucen hebras Evanescentemente finas de diseños reales y posibles: cosas que pueden *hacer cosas* aparte de tan sólo quedarse esperando el momento de desmoronarse, sujetas al mandato inexorable de la segunda ley de la termodinámica.

Sólo con esta idea bosquejada del Espacio de Diseño en la cabeza, ahora podemos ver «de una ojeada» cómo formular una serie de preguntas que han rondado el debate sobre la evolución desde tiempos de Darwin: ¿hay *algún* diseño real, natural o artificial, que no se derive en última instancia de ese único árbol de la vida, ya sea directa o indirectamente?

Algunas respuestas para considerar:

1. No.
2. Sí, algunas maravillas de la naturaleza son demasiado asombrosas, demasiado «irreductiblemente complejas» como para haber llegado mediante el proceso cuesta arriba de diseño por evolución. Un Diseñador Inteligente debe de haberlas creado de manera independiente.
3. Sí, algunos artefactos humanos, como las obras de Shakespeare y el teorema de Gödel, son demasiado maravillosos para ser «simples» productos de cerebros humanos evolucionados; son obras de *genio* milagroso y no se pueden explicar —no se pueden lograr— con el proceso cuesta arriba de diseño por evolución.

Las respuestas *2* y *3* pueden caracterizarse como profundamente dependientes del contraste entre el proceso completamente ineficiente y cuesta arriba de la selección

natural («Ignorancia Absoluta», como lo llama MacKenzie) y la ágil brillantez aparentemente espontánea de Mozart, o cualquier otro genio «divino» de tu elección (el apartado 49 explora esto). Con nuestra analogía entre el *levantamiento* y la I+D, podemos decir que cada uno proclama la necesidad de un *gancho celeste,* en inglés *skyhook,* que el *Oxford English Dictionary* define así:

> *skyhook.* Originalmente de la aeronáutica. Artilugio imaginario para unión con el cielo; medio imaginario de suspenderse del cielo.

El primer uso que registra el *Oxford English Dictionary* data de 1915: «Un piloto de aeroplano al que le ordenaron permanecer otra hora en el mismo lugar (en el aire) responde: "La máquina no está equipada con ganchos celestes"». Posiblemente el concepto del gancho celeste descienda del *deus ex machina* de la antigua dramaturgia griega: cuando los dramaturgos mediocres se daban cuenta de que sus tramas llevaban a sus héroes a dificultades inexorables, a menudo se sentían tentados a mover la manivela para que un dios entrara en escena, como Supermán, para salvar sobrenaturalmente la situación. O tal vez la evolución folclórica convergente creó completamente por su cuenta los ganchos celestes. Sería maravilloso tener ganchos celestes; serían magníficos para levantar objetos pesados y librarlos de las circunstancias difíciles, además de que acelerarían toda clase de proyectos de construcción. Es triste decirlo, pero son imposibles.[16]

Hay *grúas*, sin embargo. Las grúas pueden hacer el trabajo de levantamiento que podrían hacer nuestros ganchos celestes imaginarios, y lo hacen de una manera honesta que no da nada por sentado. Lo malo es que son caras. Tienen que diseñarse y construirse usando partes de uso corriente que ya se tienen a mano, y deben colocarse en una base firme de terreno existente. Los ganchos celestes son levantadores milagrosos, insoportados e insoportables. Las grúas son igual de buenas para levantar, y tienen la ventaja de que son reales. Cualquiera que, como yo, toda la vida haya sido un curioso que se asoma a los sitios en construcción habrá observado con cierta satisfacción que a veces basta con una pequeña grúa para levantar una gran grúa. Y a muchos otros curiosos se les habrá ocurrido que en principio esta gran grúa podría usarse para permitir o acelerar la fabricación de una grúa todavía más espectacular. Las grúas en cascada son una táctica que rara vez, o nunca, se usa en proyectos de construcción del mundo real, pero en principio no hay límite a la cantidad de grúas que podrían organizarse en serie para conseguir algun objetivo imponente.[17]

Ahora imagina todo el «levantamiento» que tiene que hacerse en el Espacio de Diseño para crear los magníficos organismos y otros artefactos que se encuentran en nuestro mundo. Desde los albores de la vida deben de haberse atravesado inmensas

distancias con las primeras y más simples entidades autorreproducibles, que se extendían hacia afuera (diversidad) y hacia arriba (excelencia). Darwin nos ha ofrecido una explicación del proceso de levantamiento más crudo, rudimentario y tonto imaginable —la cuña o plano inclinado de la selección natural—. Al dar pasos pequeños —lo más pequeños posible—, este proceso gradualmente, a lo largo de millones y millones de años, puede atravesar estas enormes distancias. Al menos eso dice Darwin. En ningún momento se necesitaría nada milagroso que viniera de las altas esferas. Cada paso se ha conseguido escalando mediante **algoritmos** brutos y mecánicos, construidos desde la base por los empeños de un ascenso anterior.

Parece increíble. ¿Podría haber pasado realmente? ¿O el proceso cada tanto (a lo mejor sólo al mero principio) necesitó que una u otra clase de gancho celeste le «echara una manita»? Durante más de un siglo, los escépticos han estado tratando de encontrar pruebas de que simplemente no hay modo de que la idea de Darwin funcione, al menos no *en todo*. Han estado deseando ganchos celestes y a la caza de ellos, como excepciones a lo que ellos consideran la sombría imagen del algoritmo de Darwin agitándolo todo. Y una y otra vez han planteado dificultades verdaderamente interesantes: saltos y lagunas, y otras maravillas que al principio en efecto parecen necesitar ganchos celestes. Sin embargo, en eso han llegado las grúas, muchas veces descubiertas por los mismos escépticos que deseaban encontrar un gancho celeste.

Llegó el momento de algunas definiciones más cuidadosas. Entendamos que un *gancho celeste* es una fuerza, poder o proceso de «primero la mente», una excepción al principio de que todo el diseño, y el diseño aparente, es en última instancia el resultado de un mecanismo ciego, carente de motivos. En contraste, una *grúa* es un subproceso, o una característica especial de un proceso de diseño del que puede demostrarse que permite la aceleración local del lento proceso básico de la selección natural, *y* del que puede demostrarse que es él mismo el producto predecible (o explicable retrospectivamente) del proceso básico. Algunas grúas son obvias y no causan polémica; otras todavía son tema de disputas muy fructíferas. Sólo para dar una impresión general de la amplitud y la aplicación del concepto, permítaseme señalar tres ejemplos muy diferentes.

La *simbiosis* es una grúa. Uno de mis ejemplos favoritos es el origen simbiótico de la célula eucarionte. Si ves el árbol de la vida (figura 13), observarás que el glorioso abanico de la vida multicelular, que incluye todas las plantas y animales, ocurre después de la evolución de los eucariontes. Durante aproximadamente 1 000 millones de años en este planeta, las únicas cosas vivientes fueron organismos unicelulares, bacterias y arqueas, que juntas se conocen como procariontes. Luego, un día de suerte, un procarionte se tropezó con otro (seguramente ocurría a menudo en aquel entonces, tal como sigue ocurriendo ahora), y en lugar de que uno de ellos envolviera y desarmara al otro (se lo

comiera), o lo ahuyentara, ambos juntaron fuerzas y formaron una clase completamente nueva de cosa viviente, con (aproximadamente) el doble de partes móviles y una mayor variedad de partes activas. Dos linajes de I+D que habían estado avanzando pesadamente cada uno por su lado durante millones de años, cada uno puliendo sus propios talentos característicos, se juntaron, y el resultado fue —resultó ser— un éxito de la noche a la mañana (la «transferencia tecnológica» no siempre ni a menudo arroja buenos resultados, pero cuando lo hace puede ser espectacular). Ninguno de los dos linajes tuvo que tomarse la molestia de reinventar lenta y pesadamente todos los trucos y sistemas del otro, y como la combinación de sus talentos resultó ser un beneficio neto (*sinergia* es la palabra de moda para esto), este *eu-carionte* («buen» + «célula») fue más apto que cualquiera por sí solo, así que el linaje de los eucariontes prosperó. La afirmación de que el origen de los eucariontes fue semejante suceso de simbiosis era polémica, pero tantas pruebas provenientes de distintas líneas de investigación la apoyan que ahora figura sólidamente en los libros de texto. (Sería gratificante informar que el origen del flagelo bacteriano, durante mucho tiempo el buque insignia entre los ejemplos de «complejidad irreductible» para la gente del diseño inteligente, también ha demostrado deberse a la grúa ayudadora de la simbiosis, y aunque la difunta Lynn Margulis, bióloga y paladín del origen simbiótico de los eucariontes, defendió vigorosamente esta afirmación, las pruebas actualmente favorecen otras explicaciones evolutivas de la fuente del flagelo. Dado que hay tanto de donde elegir, la afirmación de que los flagelos necesitan un gancho celeste ahora está muy desamparada.)

Como podemos notar al dar un vistazo al árbol de la vida, los eucariontes crearon el marco para toda la vida multicelular. Todos los linajes multicolores situados a la derecha de las bacterias y las arqueas son eucariontes. En una primera aproximación, todas las cosas vivientes que pueden verse a simple vista son eucariontes. La revolución eucarionte abrió inmensas regiones de Espacio de Diseño, pero no ocurrió *a fin de* hacer accesibles todos estos diseños. Las grúas deben «pagarse» en el acto, pues acarrean beneficios inmediatos a quienes tienen las innovaciones de diseño. Sin embargo, una vez establecidas pueden tener otros efectos profundos. (Del mismo modo, las computadoras no se inventaron *a fin de* posibilitar los procesadores de palabras y el internet, pero una vez que el espacio de las posibles aplicaciones de las computadoras se volvió accesible, los procesos de diseño trabajaron a toda máquina para crear todas las «especies» de las que dependemos todos los días.)

Hoy en día los teóricos de la evolución están generalmente de acuerdo en que *la sexualidad* es una grúa. Las especies que se reproducen sexualmente pueden moverse a lo largo del Espacio de Diseño a mayor velocidad que las especies que se reproducen asexualmente. Es más, pueden «discernir» mejoras de diseño sobre la marcha que son «invisibles» a las especies que se reproducen asexualmente.[18] Esto, sin embargo, no

puede ser la razón de ser de la sexualidad. La evolución no puede ver hacia delante, así que cualquier cosa que construya debe tener un beneficio inmediato para compensar el costo. Como recientemente han insistido algunos teóricos, la «elección» de reproducirse sexualmente acarrea un enorme costo *inmediato:* los organismos mandan sólo 50 por ciento de sus genes en cualquier transacción (por no hablar del esfuerzo y el riesgo que implica asegurar una transacción, para empezar). Así, el beneficio *de largo plazo* de las crecientes eficiencia, agudeza y velocidad del proceso de rediseño —las características que hacen de la sexualidad una grúa magnífica— no es nada para las miopes competencias locales que deben determinar qué organismos se favorecen en la generación inmediatamente posterior. Algunos otros beneficios de corto plazo deben de haber mantenido la presión selectiva positiva que se requiere para que la reproducción sexual sea una oferta que pocas otras especies pueden rechazar. Compiten entre sí una variedad de hipótesis atractivas que podrían resolver este enigma, las cuales John Maynard Smith les planteó convincentemente a los biólogos.[19] Para una lúcida introducción a este importante tema de la biología evolutiva, véase el libro de Matt Ridley *The Red Queen: Sex and the Evolution of Human Nature.*

La sexualidad revela que puede existir una grúa de gran poder que no se creó *a fin de explotar* ese poder; se creó por otras razones, aunque su poder como grúa puede ayudar a explicar por qué se ha mantenido desde entonces. Una grúa que evidentemente se creó para ser una grúa es la *ingeniería genética.* Los ingenieros genéticos ahora pueden indudablemente dar enormes saltos a través del Espacio de Diseño, creando organismos que nunca habrían evolucionado por medios «ordinarios». Esto no es ningún milagro *dado que los mismos ingenieros genéticos (y los artefactos que usan en su trabajo) son completamente productos de anteriores procesos evolutivos más lentos.* Si los creacionistas tuvieran razón en que la humanidad es una especie independiente, divina e inaccesible por los salvajes caminos darwinianos, entonces, después de todo, la ingeniería genética no sería una grúa, por haber sido creada con la ayuda de un enorme gancho celeste. No imagino que ningún ingeniero genético piense esto de sí mismo, pero esa lógica le permitiría colocarse en esa elevada, si bien precaria, posición. Esta idea es menos evidentemente tonta: si los cuerpos de los ingenieros genéticos son producto de la evolución, pero sus *mentes* pueden hacer cosas creativas que son irreductiblemente no algorítmicas o inaccesibles por cualquier camino algorítmico, entonces los saltos de la ingeniería genética podrían implicar un gancho celeste. Más adelante analizaremos brevemente esta posibilidad.

39. Competencia sin comprensión

MacKenzie describió la idea de Darwin de Ignorancia Absoluta, fuente de «todos los logros de la habilidad creativa», como una «extraña inversión del razonamiento», pues voltea patas arriba una de las ideas más «evidentes» que tenemos: la comprensión es la *fuente* de la competencia. ¿Por qué mandamos a nuestros hijos a la escuela, y por qué subrayamos los «conceptos» por encima de la «memorización»? Porque creemos que la mejor ruta para la competencia, en cualquier esfera de actividad, es la comprensión. ¡No te conformes con ser una bestia de carga! ¡*Comprende* los principios de lo que estemos haciendo para que podamos hacerlo mejor! Con toda seguridad éste es un consejo excelente en la mayor parte de las áreas de actividad humana. Reconocemos excepciones extremas, como la de los músicos dotados que «tocan de oído» y son incapaces de leer una nota, o los atletas innatos que siempre hacen maravillas pero no pueden explicar cómo o por qué las hacen, y en consecuencia no pueden entrenar a otros. Y tenemos también a las personas con el «síndrome del sabio», con extrañas combinaciones de incompetencia en casi todos los aspectos y destrezas sobrehumanas en algún dominio limitado. Sin embargo, casi siempre la regla general es difícil de negar: la comprensión normalmente es la llave a la competencia (humana).

Y Darwin realmente invierte ese razonamiento al mostrar, como MacKenzie tan vívidamente lo expresa, que la Ignorancia Absoluta es el artífice. El proceso de selección natural es extraordinariamente competente —piensa en la Segunda Ley de Orgel— pero completamente ciego y mecánico. Es más, los organismos que diseña se benefician de todo su exquisito equipo sin necesitar entender por qué o cómo son tan dotados. Mi ejemplo favorito es el cuclillo. Los cuclillos son *parásitos de puesta* que no hacen sus propios nidos. En vez de eso, el cuclillo hembra subrepticiamente pone su huevo en el nido de una pareja huésped de alguna otra especie de ave, donde recibirá las atenciones de sus involuntarios padres adoptivos. A menudo, el cuclillo hembra rodará uno de los huevos del huésped fuera del nido… ¡por si acaso los padres huéspedes saben contar! Y en cuanto el polluelo cuclillo sale del cascarón (y tiende a incubarse más rápidamente que los huevos del huésped), se toma grandes molestias para tirar del nido cualquier huevo que haya quedado ahí. ¿Por qué? Para maximizar la atención que recibirá de sus padres adoptivos.

40. Razones de libre flotación

La selección natural es un descubridor de razones automático; «descubre», «respalda» y «enfoca» razones a lo largo de muchas generaciones. Las comillas de relieve son para recordarnos que la selección natural no tiene una mente, ella misma no tiene razones, y sin embargo es competente para realizar esta «tarea» de refinamiento de diseño. Esto es un ejemplo de competencia sin comprensión. Sólo asegurémonos de saber cómo sacar provecho de las comillas de relieve. Piensa en una población en la que haya muchas variaciones. A algunos miembros les va bien (multiplicándose); a la mayoría no. En cada caso podemos preguntar *por qué*. ¿Por qué éste tuvo descendencia que sobrevivió mientras que estos otros no? En muchos casos, en la mayoría, no hay *ninguna razón en absoluto;* es pura y simple suerte, buena o mala. Pero si hay un subconjunto, a lo mejor uno muy pequeño, de casos en los que hay una respuesta, una diferencia por la que resulta que no da lo mismo, entonces lo que esos casos tienen en común proporciona el germen de una razón. Esto permite que la funcionalidad se acumule mediante un proceso que ciegamente rastrea razones y crea cosas que tienen propósitos pero no necesitan conocerlos. El principio de «información clasificada», que la reserva a quienes necesitan tenerla, reina en la biosfera, y la selección natural misma no necesita saber lo que está haciendo.

Así que había razones antes de que hubiera quienes se representaran las razones. He llamado a las razones que la evolución rastrea «razones de libre flotación», término que por lo visto les ha puesto los pelos de punta a no pocos pensadores, que sospechan que estoy conjurando alguna clase de espíritus, extrañas ideas inmateriales que no tienen nada que hacer en la sobria explicación de la realidad de un materialista. Para nada. Las razones de libre flotación no son más espirituales o problemáticas que los números o los centros de gravedad. Había nueve planetas antes de que la gente inventara maneras de expresar la aritmética, y los asteroides tenían centros de gravedad antes de que hubiera físicos a los que se le ocurriera la idea e hicieran cálculos con ella. Es un error confundir los *números* con los *numerales* (arábigos, romanos o lo que sea) que usamos para nombrarlos. Los numerales son invenciones humanas; los números no. Las razones, en el sentido en que estoy usando el término, son como los números, no como los numerales. Todos deberíamos estar encantados de hablar de las razones reveladas por la evolución antes de que investigadores humanos o cualesquiera otras mentes las expresaran o representaran. Piensa en las construcciones extraordinariamente parecidas de la figura 14.

El termitero llamado catedral y la Sagrada Familia de Gaudí tienen una forma muy parecida pero son completamente diferentes en su génesis y construcción. *Hay razones* para las estructuras y formas del termitero catedral, pero ninguna de las termitas se las representa. No hay una Termita Arquitecta que haya planeado la estructura, y ninguna

termita tiene la menor idea de por qué construyen como lo hacen. Competencia sin comprensión. También hay razones para las estructuras y formas de la obra maestra de Gaudí, pero son (principalmente) razones de Gaudí. Gaudí *tenía* razones para las formas que ordenó que se crearan; *hay* razones para las formas creadas por las termitas, pero las termitas no *tienen* esas razones. Hay razones por las que los árboles extienden sus ramas, pero en ningún sentido admisible son razones de los árboles. Las esponjas hacen cosas por razones; las bacterias hacen cosas por razones; hasta los virus hacen cosas por razones. Pero no *tienen* las razones; no necesitan tener las razones.

Hay un sinfín de razones para estos comportamientos, pero por lo general los organismos no necesitan entenderlas. Están dotados de comportamientos que han sido bien diseñados por la evolución, y son los beneficiarios de estos diseños sin necesidad de saber nada al respecto. Esta característica se ve por todas partes en la naturaleza, pero tiende a estar enmascarada por nuestra tendencia de adoptar la **postura intencional** e interpretar la conducta como más consciente y racional de lo que realmente es. ¡Qué listas son las termitas para proveer a sus termiteros de aire acondicionado mediante pozos de ventilación bien ubicados! ¡Qué sagaz la ardilla para almacenar comida para el invierno! ¡Qué astuto el lucio para acechar inmóvil mientras se aproxima su presa! Éstas son en efecto excelentes estrategias para el éxito en el implacablemente competitivo mundo de la naturaleza, pero sus beneficiarios no necesitan apreciar lo mismo que nosotros cuando las entendemos. Somos las primeras mentes que se representan las razones que explican el éxito de estos arreglos.

41. ¿Entienden los chapulines los números primos?

Sólo para reiterar lo que quiero decir al hablar de la independencia de números y numerales, las razones de libre flotación y el razonamiento representado, considera el ejemplo de los chapulines de 17 años. En 1977, Stephen Jay Gould escribió un revelador capítulo lleno de admiración por el hecho curioso de que las cigarras (igual que el «chapulín de 17 años») tienen ciclos reproductivos con duraciones iguales, en años, a números primos —13 años, o 17, pero nunca 15 o 16, por ejemplo—. «Como evolucionistas —dice— buscamos respuestas a la pregunta por qué. Por qué, en particular, debería evolucionar una sincronicidad tan asombrosa, y por qué debería ser tan largo el periodo entre episodios de reproducción sexual.»[20] La respuesta —que, en retrospectiva, tiene todo el sentido— es que, al tener un gran número primo de años entre sus apariciones, las cigarras minimizan la probabilidad de que las descubran y luego les sigan la pista como un banquete predecible los depredadores que aparecen cada dos, tres o cinco años. Si las cigarras tuvieran una periodicidad de, digamos, 16 años, serían un premio excepcional para depredadores que aparecieran cada año, pero una fuente de alimentación más confiable para los depredadores que aparecieran cada dos o cuatro años, y una apuesta con las mismas probabilidades de ganar o perder para los depredadores que entraran en fase con ellos en un calendario de ocho años. Sin embargo, si la duración de su periodo reproductivo no es múltiplo de ningún número más chico, para ninguna especie valdrá la pena «intentar» seguirles la pista… a menos que tenga la suerte de tener exactamente la misma periodicidad (o algún múltiplo de ella; el mítico comechapulines de 34 años estaría en Jauja).

Debería estar claro que la solidez de esta explicación (que posiblemente no se haya determinado) no depende de ninguna hipótesis que sugiera que los chapulines entienden aritmética, ya ni hablemos de los números primos. Tampoco depende de que el proceso de selección natural *entienda* los números primos. El proceso mecánico, carente de toda comprensión, de la selección natural puede *explotar* esta importante propiedad de algunos números sin tener en absoluto que entenderlos. Para otro ejemplo: ni las abejas ni la Madre Naturaleza necesitan *entender* la geometría que declara el hexágono como la forma ideal para las celdillas de un panal. Podrían citarse muchos ejemplos más de la competencia matemática, carente de comprensión, del proceso evolutivo.

42. Cómo explicar los saltos de las gacelas

¿Recuerdas la pasmosa regularidad que se reveló en la bomba de intuición sobre las **dos cajas negras**? No necesitabas tener una interpretación semántica o intencional de las cajas para verla: las α causaban rojo y las β causaban verde —una regularidad evidente que necesitaba explicación—. Los científicos entendían exactamente cómo funcionaba cada paso en el proceso de cómputo y transmisión que se daba cada vez que se apretaba un botón, pero no podían explicar la generalización. Sí se necesita una interpretación semántica para explicar por qué existe la regularidad. En otras palabras, el nivel «macrocausal» en el que se expresa la explicación no se «reduce» al nivel «microcausal».

Aquí hay otro ejemplo de este fenómeno de lo más general. Probablemente habrás visto filmaciones de gacelas a las que un depredador persigue a lo largo de la llanura, y habrás observado que algunas de ellas dan unos brincos muy altos y característicos en sus intentos de escapar de su perseguidor (a esto en inglés se le llama *stotting*). ¿Por qué dan esos saltos? Son claramente benéficos, porque a las que los dan rara vez las atrapan y se las comen. Ésta es una regularidad causal que se ha observado cuidadosamente, igual que la regularidad entre botones y luces. Y también ésta puede dejar perplejo. Ninguna explicación de las acciones de todas las proteínas en todas las células de todas las gacelas y de los depredadores que las persiguen podría revelar por qué existe esta regularidad. Para eso necesitamos la rama de la teoría evolutiva conocida como la teoría de las señales costosas.[21] Las gacelas más fuertes y veloces saltan para anunciarle su buena forma a su perseguidor, en efecto, mandando la señal «Ni te molestes en perseguirme; soy muy difícil de atrapar: mejor concéntrate en alguno de mis primos que no pueden saltar como yo: son una comida mucho más sencilla», y el perseguidor interpreta esto como una señal honesta, difícil de fingir, e ignora a la que da saltos. Esto son las razones de libre flotación, y ni la gacela ni el depredador tienen necesidad de apreciarlas. Es decir, la gacela puede ser completamente ignorante de por qué es buena idea saltar si puede, y el depredador, digamos un león, puede no entender por qué las gacelas que saltan son una presa relativamente poco atractiva, pero si las señales no fueran honestas y costosas, no podría persistir en la carrera armamentística evolutiva entre depredador y presa (si la evolución tratara de usar una señal «barata», como un coletazo, que todas las gacelas, sin importar qué tan fuertes o débiles fueran, pudieran mandar, para los leones no valdría la pena prestarle ninguna atención, así que no lo harían). Estas explicaciones basadas en las razones de libre flotación no son reducibles a explicaciones de niveles inferiores, como el molecular, pero es importante reconocer que aunque la explicación de por qué y cómo funciona el salto de las gacelas está hecha desde la **postura intencional** (en cuanto a qué sería racional que concluyera un león a partir del salto), el león o la gacela no necesitan entender el significado del salto para que

funcione; sólo necesitan **maomeno** entenderlo.

43. Cuidado con el Mamífero Primario

Puedes pensar que eres un mamífero, y que perros, vacas y ballenas son mamíferos, pero en realidad no existen los mamíferos. ¡No podrían! Aquí hay un argumento filosófico que lo demuestra:[22]

1. Todos los mamíferos tienen una madre que es mamífera.
2. Si ha habido mamíferos, ha habido sólo un número finito de mamíferos.
3. Pero si ha habido incluso un mamífero, entonces por *1* ha habido un número infinito de mamíferos, que contradice *2,* así que no puede haber habido ningún mamífero. Es una contradicción en sus términos.

Como sabemos perfectamente bien que hay mamíferos, nos tomamos en serio este argumento sólo como incentivo para descubrir la falacia que acecha dentro de él. Es una franca **reducción al absurdo**, así que hay que ceder en algo. Y sabemos, de manera general, en qué ceder: si te remontas lo suficientemente atrás en el árbol genealógico de cualquier mamífero, tarde o temprano llegarás a los *terápsidos,* extraña y extinta especie puente entre los reptiles y los mamíferos. Ocurrió una transición gradual entre el reptil y el mamífero claramente diferenciados, con muchos intermediarios difíciles de clasificar llenando los huecos. ¿Qué deberíamos hacer para trazar las líneas a lo largo de este espectro de cambio gradual? ¿Podemos identificar un mamífero, el Mamífero Primario, que no tuviera una madre que fuera mamífera, y entonces negar la premisa 1? ¿Sobre qué base? Sea cual sea la base, competirá con la base que pudiéramos usar para apoyar el veredicto de que ese animal *no* fuera un mamífero —después de todo, su madre era un terápsido—. ¿Qué mejor prueba de *terapsidad* puede haber? Supón que hacemos una lista de 10 grandes diferencias que sirvieran para distinguir a los terápsidos de los mamíferos, y declaramos que tener cinco o más señales de mamífero hace que un animal sea un mamífero. Aparte de ser arbitraria —¿por qué 10 diferencias, y no seis o 20?, ¿y no deberían estar en orden de importancia?—, cualquier línea divisoria como ésa generará grandes cantidades de veredictos no deseados, pues durante el larguísimo periodo de transición entre los terápsidos obvios y los mamíferos obvios habrá muchos casos en los que los mamíferos (en virtud de nuestra regla de cinco o más) se hubieran apareado con terápsidos (menos de cinco señales de mamífero) y tuvieron vástagos que fueron terápsidos nacidos de mamíferos, mamíferos nacidos de terápsidos nacidos de mamíferos, etc. Desde luego, necesitaríamos una máquina del tiempo para ver todas estas anomalías, pues no se pueden apreciar los detalles después de todos esos millones de años. Da lo mismo, pues los detalles realmente no importan a la larga. ¿Qué debemos hacer? *Debemos sofocar nuestro deseo de trazar líneas.* No necesitamos trazar líneas. Podemos vivir con el nada escandaloso ni misterioso hecho de que todos estos cambios

graduales se acumularon a lo largo de muchos millones de años y a la larga produjeron indudables mamíferos. Del mismo modo, nadie necesita calibrar las diferencias entre lagos y lagunas o humedales y pantanos, ni siquiera los limnólogos (quienes estudian las aguas continentales).

Los filósofos, sin embargo, tienden a ser prolijos y quisquillosos usuarios de palabras. Desde que Sócrates insistió en pedir que se le dijera exactamente qué marcaba las características definitorias de la virtud, el conocimiento, la valentía y demás, los filósofos han estado tentados por la idea de detener la amenaza de una regresión infinita como ésta identificando algo que es —debe ser— *el* tapón de la regresión: el Mamífero Primario, en este caso. A menudo los hace aterrizar en doctrinas que se regodean en el misterio, o al menos en el desconcierto, y los compromete con el *esencialismo.* El Mamífero Primario debe de ser cualquier mamífero en el conjunto de mamíferos que fue el primero en tener todas las características *esenciales* de los mamíferos. Si no hay una esencia definible de *mamífero* —y la biología evolutiva nos muestra que no hay tales esencias—, estos filósofos están en problemas. Así, como regla general, plantéate ignorar la exigencia de los filósofos por una esencia, una característica definitoria, un «hacedor de verdad». Normalmente —no siempre— da inicio a una persecución inútil que puede ser entretenida pero a lo mucho moderadamente ilustradora.

Abandonar esta exigencia es un trago verdaderamente amargo para muchos filósofos. Los métodos racionales para hacer filosofía que se han transmitido desde Sócrates casi siempre piden «definir tus términos» de un modo que compromete a todos los participantes en la discusión con una doctrina de esencialismo, aunque sólo sea *por mor del argumento,* como decimos. Si tenemos que abandonar el esencialismo, algunas de nuestras formas argumentativas predilectas se vuelven casi inservibles. Considera, por ejemplo, la estructura de los argumentos que empiezan con una *disyunción* en apariencia trivialmente verdadera:

O bien A o bien no A (¿cómo puedes poner *eso* en duda?)
Si eliges el camino A, entonces blablablá, así que llegas a la conclusión C;
y si empiezas con no A, entonces blablablá, ¡*también* llegas a C!
Por lo tanto se demuestra C.

¿Y qué si al parecer hay muchos casos intermedios en los que no está claro si son A o no A (mamíferos o no mamíferos, vivos o no vivos, conscientes o no conscientes, una creencia o no una creencia, moral o no moral, etc.)? Para dejar a un lado esta preocupación, tienes que «trazar una línea» que distinga la A de la no A y destierre todo lo que hemos dicho sobre lo maomeno. Sin esa nítida frontera que señale la esencia de lo que sea que esté en juego, el argumento simplemente no puede elaborarse. Esos

argumentos funcionan de maravilla en matemáticas, donde realmente *puedes* trazar las líneas. Todos los enteros realmente son pares o nones, y todo número realmente es racional o irracional, y todo polígono realmente es un triángulo (de tres lados) o no lo es. Fuera de esos dominios abstractos, estos argumentos dejan de funcionar tan bien.

La insistencia en que debe de haber un Mamífero Primario, aunque nunca sepamos cuándo y dónde existió, es un ejemplo de *realismo histérico.* Nos invita a reflexionar en que si tan sólo supiéramos lo suficiente, veríamos —*tendríamos* que ver— que hay una propiedad especial de la *mamiferidad* que define a los mamíferos de una vez y para siempre. Negarlo, dicen a veces los filósofos, es confundir metafísica con epistemología: el estudio de lo que (realmente) *es* con el estudio de lo que podemos *saber* acerca de lo que es. Yo respondo que puede haber ocasiones en que los pensadores sí se descarrilan al confundir una pregunta metafísica con una pregunta (meramente) epistemológica, pero esto debe demostrarse, no sólo afirmarse.

44. ¿Cuándo ocurre la especiación?

Una característica curiosa de la evolución por selección natural es que depende fundamentalmente de sucesos que «casi nunca» ocurren. Por ejemplo, la especiación, el proceso en el que una nueva especie se genera por alejarse de su especie progenitora, es un suceso excesivamente raro, pero cada uno de los millones de especies que han existido en este planeta empezó con un suceso de especiación. Todos los nacimientos de todos los linajes son un potencial suceso de especiación, pero la especiación casi nunca ocurre, ni una vez en un millón de nacimientos. La mutación del ADN casi nunca ocurre —ni una vez en un billón de copias—, pero la evolución depende de ella. Es más, la inmensa mayoría de las mutaciones son o bien perjudiciales o bien neutrales. Una mutación fortuitamente «buena» casi nunca ocurre. Sin embargo, la evolución depende de esos sucesos, casi los más raros entre los raros.

Considera una bomba de intuición sobre una extraordinaria posibilidad: en este momento, hasta donde sabemos, solamente hay una especie de homínidos en el planeta, los *Homo sapiens*. Pero supón que dentro de 50 años todos nuestros descendientes, excepto un afortunado puñado de ellos, son exterminados por un virus que deja sólo dos grupos de sobrevivientes: 1 000 esquimales que viven en la lejana isla de Cornwallis, frente a Groenlandia, y 1 000 andamaneses que viven en un maravilloso aislamiento en unas islas en medio del océano Índico. Estas dos poblaciones han estado separadas miles de años y han desarrollado diferencias fisiológicas características en respuesta a sus muy diferentes medios, pero no tenemos ninguna buena razón para cuestionar la suposición estándar de que son miembros de nuestra especie. Supón que estas poblaciones siguen aisladas una de otra tanto geográfica como reproductivamente otros 10 000 años, y que a la larga vuelven a poblar el globo con dos especies —que es lo que descubren cuando finalmente se encuentran y se dan cuenta de que no tienen ningún interés en aparearse y además sus pocos inadvertidos intentos de aparearse son infructuosos, una señal estándar de *especiación alopátrica,* aislamiento reproductivo que se da con el tiempo por el aislamiento geográfico—. Podrían preguntarse exactamente cuándo ocurrió la especiación. Es probable que su último ancestro común haya vivido hace más de 30 000 años, pero la especiación no ocurrió ahí y en ese momento (y hasta donde sabemos, todavía no ocurre), pero aún sería posible que en unos cuantos miles de años se revelara que la especiación había ocurrido un tiempo antes de que las dos poblaciones se reunieran. ¿Ocurrió la especiación antes de los albores de la agricultura o después de la creación de internet? No hay una respuesta no arbitraria que pudiéramos defender cómodamente. Debe de haber habido un último ancestro común, o *concestro* —para usar el término de Dawkins—, que vivió tal vez hace 30 000 años, y pudo ser que a la larga los descendientes de este individuo *hayan sido* los fundadores de dos especies diferentes, pero hoy en día todavía no se determina, de un modo o del otro, si el suceso de

especiación empezó entonces.[23]

Aquí tendríamos un suceso, un nacimiento que pudiera haber tenido un papel fundamental en la historia humana (y posthumana), un suceso que ocurrió en un momento y un lugar precisos pero que no adquirió su estatus especial hasta que milenios de *secuelas* fijaron ese papel, que nunca fue un resultado inevitable. Todo lo que se necesitaría para impedir que ese nacimiento fuera un suceso de especiación sería un solo barco (o un avión) cargado de isleños haciendo un viaje que llevara a la reunificación «prematura» de las ramas. Uno puede *imaginar* que la especiación de hecho ocurrió en algún momento preciso pero incognoscible en el intervalo entre el aislamiento inicial de los linajes y la demostración final de que se trata de dos especies, ¿pero cómo podría definirse ese momento crítico? Quiza como la primera vez en que, gracias a la acumulación de divergencia cromosómica entre los dos linajes, *si* se intentó algún apareamiento interlinaje *habría* resultado ser infértil. Pero las conjeturas sobre tales casos infundados tienen escasa importancia.

La construcción del ferrocarril transcontinental dividió las manadas de bisontes americanos en poblaciones reproductivamente aisladas, pero Buffalo Bill Cody y sus secuaces cortaron de raíz ese potencial suceso de especiación en un abrir y cerrar de ojos por haber prácticamente exterminado a todas las poblaciones excepto una. A menudo los sucesos ambientales dividen las poblaciones de conespecíficos en dos (o más) grupos aislados, que permanecen reproductivamente aislados uno de otro durante unas cuantas generaciones, y casi siempre ocurre que o bien estos grupos se vuelven a reunir o uno de ellos se extingue. Así, aunque esos primeros pasos hacia la especiación deben de ocurrir con bastante frecuencia, casi nunca llevan a la especiación, y cuando lo hacen, es un resultado que toma muchos cientos de generaciones establecer. Absolutamente nada de las circunstancias de la separación inicial podría indicarte si se trataba del principio de un suceso de especiación, aunque tuvieras un conocimiento físico perfecto de todas las moléculas del mundo en ese momento. El mismo concepto de especie es un **maomeno** concepto. A los perros domésticos, los coyotes y los lobos se les llama especies, y sin embargo las descendencias de coyotes con perros y perros con lobos son bastante comunes, así que a lo mejor deberíamos considerarlos «oficialmente» como tres simples variantes, ni siquiera subespecies, de *Canidae.* Las generalizaciones sobre la posibilidad de hibridación —crear descendencia de padres de diferentes especies— son difíciles de demostrar, lo que no sorprende si piensas que todos los individuos de todas las especies presentan diferencias sutiles de todos los otros individuos de su especie. Esto no preocupa especialmente a los biólogos: han aprendido a no preocuparse por definiciones o esencias, pues los procesos que crean todos los casos intermedios se entienden bien.

45. Hacedores de viudas, Eva Mitocondrial y coronaciones retroactivas

Una mujer en Nueva York puede de pronto adquirir la propiedad de ser una viuda en virtud de los efectos que una bala acaba de tener en el cerebro de algún hombre en Dodge City, a más de 1 500 kilómetros de ahí. (En tiempos del Viejo Oeste había un revólver apodado el Hacedor de Viudas. Si un revólver particular hizo honor a su sobrenombre en determinada ocasión es algo que no podría resolverse ni con el examen más exhaustivo de la escena del crimen.) El ejemplo obtiene su curiosa capacidad de saltar a través del tiempo y el espacio dada la naturaleza convencional de la relación de matrimonio, en la que se considera que un suceso histórico pasado, un casamiento, crea una relación permanente —una relación formal, no una relación causal— de interés a pesar de andanzas subsecuentes e infortunios concretos (la pérdida accidental de un anillo o la destrucción del acta de matrimonio, por ejemplo).

El sistema de reproducción genética es natural, no convencional, pero funciona como un reloj, y justamente su sistematicidad nos permite pensar *formalmente* sobre cadenas causales que se extienden a lo largo de millones de años, y que de otro modo sería prácticamente imposible designar, referirse a ellas o rastrearlas. Esto nos permite interesarnos en relaciones todavía más distantes y localmente invisibles que la relación formal del matrimonio, así como razonar sobre ellas con rigor. La especiación, como el matrimonio, es un concepto anclado en un sistema de pensamiento estrecho, que se puede definir formalmente, pero, a diferencia del matrimonio, no tiene hitos convencionales —bodas, anillos, actas— con los cuales pueda observarse. Como acabamos de ver, la especiación también es un fenómeno curiosamente de «larga distancia», tanto en espacio como en tiempo. Las especies son colecciones de bordes confusos continuadas en el tiempo, y sólo retrospectivamente (y arbitrariamente) podríamos identificar en tal muchedumbre algún organismo individual como el Mamífero Primario (así que ni te molestes). Podremos apreciar mejor esta característica de la especiación si primero observamos otro caso de coronación retroactiva, la concesión del título de Eva Mitocondrial, que no es arbitrario.

Los organismos individuales tienen lindes un poco más nítidos y en consecuencia identidades más claras que las especies, pero incluso entre ellos hay muchos que nadan entre dos aguas. Tomemos el ejemplo más sorprendente: de los aproximadamente 10 000 billones (más o menos) de células metidas adentro de tu ropa, ¡nueve de cada 10 no son humanas! Así es: los visitantes simbiontes de miles de especies superan en número a tus células huésped —las células que descendieron del cigoto formado por la unión de tus padres—. No sólo hay bacterias sino también eucariontes, tanto microbios unicelulares como organismos multicelulares: hongos, ácaros en tus pestañas y otras partes, gusanos microscópicos y más grandes, y quién sabe cuántas cosas más. Eres un ecosistema

ambulante, y si bien algunos de los visitantes son indeseados (los hongos que causan el pie de atleta o las bacterias que provocan mal aliento o se aglomeran en torno de cualquier infección), otros son tan fundamentales que si consiguieras desalojar a todos los intrusos, morirías. Como estas células simbióticas generalmente son mucho más chicas que tus células humanas, por *peso* eres sobre todo humano, pero su peso combinado tampoco es insignificante: probablemente algunos kilos, hasta unos cinco quizá. Y luego están los virus, en cantidades todavía mayores.

Con todo, a pesar de tus límites porosos, eres, al igual que otros organismos individuales, inmediatamente distinguible de los demás, y a veces podemos señalar particularmente a un organismo individual por un papel específico desempeñado en la historia evolutiva. La Eva Mitocondrial es uno de los más famosos. Se trata de la mujer que es el ancestro directo *más reciente,* en la línea materna, de todos los seres humanos *que están vivos el día de hoy.* Todos tenemos mitocondrias en nuestras células, y nos llegan sólo a través de la línea materna, así que todas las mitocondrias en todas las células de las personas que viven en la actualidad son descendientes directos de la mitocondria de las células de una mujer determinada, a la que Rebecca Cann, Mark Stoneking y Allan Wilson llamaron Eva Mitocondrial.[24]

Las mitocondrias son pequeños organelos de nuestras células que realizan una función central en el metabolismo al captar energía de la comida para ser empleada en todos los proyectos del cuerpo, y las mitocondrias tienen su propio ADN, un rastro revelador de su origen simbiótico hace varios miles de millones de años. Analizando el orden del ADN mitocondrial de las diferentes personas hoy vivas, los científicos han podido deducir en líneas generales cuán recientemente vivió la Eva Mitocondrial, y hasta dónde vivió. Los primeros cálculos sugerían que vivió en África hace como 300 000 años, pero recientemente se ha afinado el número: vivió (casi con certeza en África) hace apenas 200 000 años. Deducir *dónde* y *cuándo* es una tarea mucho más complicada que deducir *que* hubo una Eva Mitocondrial, algo de lo que ningún biólogo duda. Considera algunas de las cosas que ya sabemos de ella, poniendo de lado las polémicas. Sabemos que tuvo al menos dos hijas que a su vez tuvieron dos hijas que sobrevivieron (si hubiera tenido sólo una hija, su hija llevaría la corona de Eva Mitocondrial). Para distinguir su título de su nombre propio, llamémosla Amy. Amy lleva el título de Eva Mitocondrial, es decir, simplemente resulta ser la fundadora materna del linaje humano actual. Es importante tener presente que probablemente *en ningún otro aspecto* había nada sobresaliente o especial en la Eva Mitocondrial; ciertamente no fue la primera mujer, ni la fundadora de la especie *Homo sapiens.* Muchas mujeres anteriores fueron incuestionablemente de nuestra especie, pero da la casualidad de que ninguna de ellas fue la fuente más reciente de todas las mitocondrias de la gente viva hoy. También es cierto que aunque la Eva Mitocondrial tuvo hijas y nietas, probablemente no era mucho más

fuerte, rápida, bella ni fecunda que todas las mujeres de su tiempo.

Para resaltar cuán poco especial era probablemente la Eva Mitocondrial —es decir, Amy—, supongamos que mañana, miles de generaciones después, una nueva enfermedad virulenta se extendiera a todo lo ancho de la tierra y en pocos años exterminara a 99 por ciento de la raza humana. Los sobrevivientes, con la suerte de haber tenido alguna resistencia innata al virus de la enfermedad, probablemente tendrían algún cercano parentesco. *Su* ancestra directa en común más cercana —llamémosla Betty — sería alguna mujer que vivió cientos o miles de generaciones después que Amy, y la corona de Eva Mitocondrial pasaría a ella, retroactivamente. Ella pudo haber sido la fuente de la mutación que siglos después se convirtió en salvadora de la especie, pero probablemente *a ella* no le sirvió de nada, pues el virus contra el cual triunfa quizá no existía en su forma virulenta en aquel entonces. El caso es que la Eva Mitocondrial sólo puede ser coronada *retroactivamente*. Este papel histórico fundamental se determina no sólo por los accidentes de la época de Amy, sino también por los accidentes de tiempos posteriores. ¡Eso sí que es una enorme contingencia! Si el tío de Amy no la hubiera salvado de ahogarse cuando tenía tres años, ¡ninguno de *nosotros* (con nuestro particular ADN mitocondrial, gracias a Amy en última instancia) habría jamás existido! Si todas las nietas de Amy hubieran muerto de hambre en la primera infancia —como les pasaba a muchos menores de tres años en aquellos tiempos—, también nosotros estaríamos sepultados en el olvido.

El mismo argumento lógico demuestra que también hay —debe haber— un Adán: el ancestro varón directo más cercano de todos los hombres y niños vivos hoy. Lo podemos llamar Adán Y-cromosómico, pues todos los cromosomas Y nos vienen de la línea paterna tal como la mitocondria nos viene de la línea materna.[25] ¿Fue el Adán Y-cromosómico el esposo o el amante de la Eva Mitocondrial? No hay manera. Como la paternidad es un asunto que consume mucho menos tiempo y energía que la maternidad, lo que es *lógicamente* posible es que el Adán Y-cromosómico haya vivido *muy* recientemente, y haya estado muy pero muy ocupado en la recámara —y junto a él Errol Flynn se quedaría corto—. Si el hombre más viejo vivo hoy tiene, digamos, 110 años, es *lógicamente* posible que el Adán Y-cromosómico fuera su padre, un donjuán de principios del siglo XX que también es el padre, abuelo o bisabuelo, y así sucesivamente, de los hombres vivientes de menor edad. A fin de cuentas, los hombres producimos miles de millones de espermatozoides, cientos de millones en cada eyaculación, así que al Adán Y-cromosómico no le tomaría más de una semana producir suficiente esperma para engendrar a toda la raza humana (¡en principio!). Pero si contamos todas las diferencias genéticas en los cromosomas Y de los hombres en todo el mundo, y calculamos cuánto tiempo tomaría para que se acumularan todas esas mutaciones, podemos concluir que la fecha real para el Adán Y-cromosómico fue hace poco menos de 100 000 años. Y una

vez más, si una epidemia arrasara con, digamos, la mitad de la población masculina, con toda probabilidad la corona de Adán Y-cromosómico pasaría a un progenitor mucho más reciente.[26]

Un hecho curioso acerca de todos los organismos individuales —por ejemplo tú, yo, tu perro o tu geranio— es que es un fundador *potencial* de una nueva especie, el primero de una larga línea de veteasaberqués, pero vendrán cientos o miles de generaciones antes de que los veteasaberqués resalten entre la multitud lo suficiente para ser reconocidos como una especie, así que la coronación tendría que ocurrir mucho después de que tú, yo, tu perro o tu geranio nos hubiéramos convertido en polvo. Tus padres podrían, entonces, resultar ser los más recientes ancestros en común de todos los miembros de dos especies humanoides… pero espera sentado. Los perros chihuahueños y los gran danés son miembros de la misma especie, *Canis familiaris,* pero si la civilización se colapsa y sus respectivos vástagos regresan al estado salvaje, ellos podrían conseguir más rápido el retraimiento de la especiación que, por ejemplo, los beagles y los basset hounds, pues, sin ayuda humana, la fertilización de un perro chihuahueño por un gran danés, o viceversa, es poco factible. Sin embargo, con toda probabilidad, antes de que eso pasara ambos linajes se extinguirían, como les ha pasado a la mayoría de los linajes a lo largo de millones de años.

Se calcula que mucho más de 99 por ciento de todos los organismos que han existido hasta hoy murieron sin tener descendencia. Sin embargo, hete aquí: de todos tus miles de millones de ancestros a lo largo de los años, de células unicelulares a gusanos, a peces, a reptiles, a mamíferos y a primates, ni uno solo de ellos murió sin haber tenido hijos. ¡Qué suerte tienes! Desde luego, cada brizna de hierba ha tenido una ascendencia igualmente larga e imponente (y cada mosquito, cada elefante y cada margarita…).

46. Ciclos

Todo el mundo sabe de los conocidos ciclos de la naturaleza a gran escala: el día sigue a la noche y la noche sigue al día; verano-otoño-invierno-primavera-verano-otoño-invierno-primavera; el ciclo del agua de evaporación y precipitación que rellena nuestros lagos, limpia nuestros ríos y restablece el suministro de agua de todas las cosas vivientes del planeta. Pero no todo el mundo aprecia cómo los ciclos —en toda escala espacial y temporal, desde la atómica hasta la astronómica— son, de manera bastante literal, los motores encendidos ocultos que hacen funcionar a todos los fenómenos maravillosos de la naturaleza. Nikolaus Otto construyó y vendió el primer motor de combustión interna de gasolina en 1861, y Rudolf Diesel construyó su motor en 1897, dos brillantes inventos que cambiaron el mundo. Cada uno de ellos explota un ciclo, el ciclo de cuatro tiempos de Otto y el ciclo de dos tiempos de Diesel, que realiza algún trabajo y luego restablece el sistema a la posición original y lo deja listo para realizar más trabajos. Los detalles de estos ciclos son ingeniosos; los ha descubierto y optimizado un ciclo de I+D de invención que data de siglos atrás. En 1937 Hans Krebs descubrió el llamado ciclo de Krebs, un motor microminiaturizado aún más elegante, pero fue inventado durante millones de años de evolución en los albores de la vida. Es la reacción química de ocho tiempos que convierte combustible —comida— en energía en el proceso de metabolismo esencial para toda vida, desde las bacterias hasta las secuoyas.

Ciclos bioquímicos como el de Krebs son responsables de todo movimiento, crecimiento, autorreparación y reproducción en el mundo viviente: ruedas dentro de ruedas dentro de ruedas, un mecanismo de relojería con billones de partes móviles, y cada reloj necesita que se le dé cuerda y se le restablezca al paso uno para que de nuevo pueda cumplir con su deber. Todos estos ciclos se han visto optimizados por el gran ciclo darwiniano de reproducción, generación tras generación, y ha conseguido mejoras fortuitas a lo largo de millones y millones de años.

En una escala completamente distinta, nuestros ancestros descubrieron la eficacia de los ciclos en uno de los grandes avances de la prehistoria humana: el papel de la repetición en la manufactura. Toma un palo y frótalo con una piedra, y no pasa nada: unos cuantos rasguños son los únicos signos visibles del cambio. Devuelve el palo al punto inicial y hazlo otra vez. Sigues sin tener prácticamente nada que pruebe tus esfuerzos. Frótalo 100 veces y sigue sin notarse casi nada. Pero hazlo unos cuantos miles de veces y podrás convertirlo en un astil de flecha asombrosamente recto. Por la acumulación de incrementos imperceptibles, el proceso cíclico crea algo completamente nuevo. La misma combinación de previsión y autocontrol que se requiere para estos proyectos fue de suyo una novedad, una inmensa mejoría frente a los procesos de construcción y formación, repetitivos pero en su mayor parte instintivos y mecánicos, de otros animales. Y esa novedad misma fue un producto del ciclo darwiniano, que en algún

momento mejoró con el ciclo, más veloz, de la evolución cultural, en la que la reproducción de la técnica no se pasó a la descendencia a través de los genes, sino que se transmitió entre individuos no emparentados entre sí que aprendieron el truco de imitar.

El primer ancestro que pulió una roca para convertirla en un hacha de mano magníficamente simétrica debe de haberse visto muy estúpido en el proceso, ahí sentado, frotando y frotando durante horas, aparentemente sin conseguir nada. Sin embargo, oculto en los intersticios de toda la repetición mecánica había un proceso de refinamiento gradual que era prácticamente invisible para un ojo que la evolución diseñó para detectar cambios que ocurren a un ritmo mucho más veloz.[27] En ocasiones, esa misma apariencia de futilidad ha inducido a error a biólogos sofisticados. En su elegante libro *Wetware,* el biólogo celular y molecular Dennis Bray describe los ciclos del sistema nervioso:

> En una vía de señalización típica, las proteínas continuamente se modifican y desmodifican. Las cinasas y los fosfatos trabajan incesantemente como hormigas en el hormiguero, añadiendo grupos de fosfato a las proteínas y luego quitándolos. Parece un ejercicio sin sentido, especialmente cuando piensas que cada ciclo de añadir y quitar le cuesta a la célula una molécula de atp —una unidad de energía preciosa—. De hecho, las reacciones cíclicas de este tipo fueron en un principio llamadas «vanas», pero el adjetivo es equívoco. La adición de grupos de fosfato a proteínas es la reacción más común en las células y sustenta una gran proporción de los cómputos que realizan. Lejos de ser vana, esta reacción cíclica le proporciona a la célula un recurso fundamental: un dispositivo flexible y rápidamente afinable.[28]

La palabra «cómputos» es una elección acertada. Los programas de computadora han estado explorando el espacio de cómputos posibles durante menos de un siglo, pero su cosecha de invención y descubrimiento hasta ahora incluye millones de lazos dentro de lazos dentro de lazos. Resulta que toda la «magia» de la cognición depende, como la vida misma, de ciclos dentro de ciclos de procesos que transforman la información, recurrentes, «reentrantes», reflexivos, que van de los ciclos bioquímicos a nanoescala dentro de cada neurona, pasando por los ciclos de generar y probar el cifrado de los sistemas perceptuales, hasta todo el ciclo de sueño del cerebro, ondas de actividad y recuperación cerebral a gran escala que los registros de los electroencefalogramas revelan.[29] El ingrediente secreto de la mejora en todos los aspectos de la vida es siempre el mismo: práctica, práctica, práctica.

Es útil recordar que la evolución darwiniana es sólo una clase de ciclo acumulativo y mejorable. Hay muchos otros. El problema del origen de la vida puede plantearse de modo que parezca irresoluble («irreductiblemente complejo»)[30] si uno argumenta, como han hecho los defensores del diseño inteligente, que como la evolución por selección

natural depende de la reproducción, no puede haber una solución *darwiniana* al problema de cómo llegó a existir la primera cosa viviente y reproductora: con toda seguridad fue increíblemente complicado, así como hermosamente diseñado —tuvo que haber sido un milagro—. Si nos da por pensar en el mundo prebiótico y prerreproductivo como una especie de caos anodino de sustancias químicas (como las partes desperdigadas del conocido avión de pasajeros armado por un ventarrón que los creacionistas nos invitan a imaginar), el problema sí parece abrumador o algo peor, pero si recordamos que el proceso clave de la evolución es la repetición cíclica (de la cual la reproducción genética no es más que un ejemplo altamente refinado y optimizado), podemos empezar a ver el camino para convertir el misterio en un enigma: ¿cómo fue que todos esos ciclos estacionales, los ciclos hidrológicos, geológicos y químicos, dando vueltas millones de años, gradualmente acumularon las condiciones necesarias para inaugurar los ciclos biológicos? Probablemente los primeros mil «intentos» fueron vanos, casi aciertos. Pero como nos recuerda la maravillosamente sensual canción de George Gershwin y Buddy DeSylva, ve qué pasa si lo «haces otra vez» (y otra y otra).[31]

Entonces una buena regla general, cuando nos enfrentamos a la aparente magia del mundo de la vida y de la mente, es buscar los ciclos que están haciendo todo el trabajo duro.

47. ¿Qué *le dice* el ojo de la rana al cerebro de la rana?

Uno de los primeros clásicos de la ciencia cognitiva es el famoso artículo «What the Frog's Eye Tells the Frog's Brain» («Lo que le dice el ojo de la rana al cerebro de la rana»), de J. Y. Lettvin y colegas.[32] En él se mostraba que el sistema visual de la rana puede percibir manchitas negras que se mueven en la retina, pequeñas sombras que en casi todas las circunstancias naturales proyectan las moscas que vuelan en los alrededores. Este mecanismo «detector de moscas» está debidamente comunicado con el sensible disparador de la lengua de la rana, lo que fácilmente explica cómo se alimentan las ranas en un mundo cruel y de ese modo ayudan a propagar su especie. Ahora bien, ¿qué *le dice* el ojo de la rana al cerebro de la rana? ¿Que allá fuera hay una mosca, o que hay una mosca-o-mosca-falsa (algún tipo de imitación de mosca), o una cosa de la clase *K* (cualquier clase de cosa que confiablemente detone este artilugio visual —recuerda el tragacuartos—)? Los teóricos del significado darwinianos (como Ruth Millikan, David Israel y yo) hemos discutido este mismísimo ejemplo, y el archicrítico de la teoría evolucionista, Jerry Fodor, se ha abalanzado para mostrar lo que, a su entender, falla con cualquier explicación evolucionista de tales significados: son demasiado indeterminados. Dichos significados no distinguen, como deberían, entre esos informes del ojo de la rana como «mosca aquí ahora» y «mosca o pequeño proyectil oscuro aquí ahora», y así sucesivamente. Pero esto es falso. Podemos usar el medio ambiente de selección de la rana (en la medida en que podamos determinar cuál ha sido) para distinguir entre los diversos candidatos. Para esto usamos exactamente las mismas consideraciones que sirvieron para resolver las preguntas sobre el significado del estado en el tragacuartos. Y en la medida en que no hay nada en el medio selectivo que de manera unívoca señale una clase particular de ocasiones, tampoco hay una verdad acerca de lo que el informe del ojo de la rana *realmente* significa. Esto puede quedar más claro si mandamos a la rana a Panamá —o, más precisamente, a un nuevo medio selectivo—.

Supongamos que los científicos reunieran una pequeña población de ranas de alguna especie atrapamoscas al borde de la extinción y para protegerlas las retuvieran en un nuevo medio: un zoológico especial para ranas en el que no hay moscas, sino más bien guardias que periódicamente lanzan bolitas de comida por delante de las ranas a su cuidado. Para regocijo del cuidador, el sistema funciona; las ranas crecen fuertes y sanas lanzando la lengua para atrapar las bolitas, y después de un tiempo hay una multitud de ranas descendientes que nunca han visto una mosca, sólo bolas de comida. ¿Qué les dicen *sus* ojos a *sus* cerebros? Si insistes en que el significado no ha cambiado, estás en apuros. El ejemplo de las ranas no es sino un caso artificialmente claro de lo que pasa todo el tiempo con la selección natural: la *exaptación* —la reutilización de una estructura

existente para una nueva función—. Como Darwin tuvo el cuidado de recordarnos, la reutilización de maquinaria para nuevos propósitos es uno de los secretos del éxito de la Madre Naturaleza. Podemos remacharlo, para quien todavía no esté convencido, suponiendo que no a todas las ranas cautivas les va igual de bien. Variaciones en la habilidad de detectar bolitas en sus sistemas visuales dan lugar a que algunas coman con menos ganas que otras, y que como resultado dejen menos progenie. En breve tendrá que haber una indudable selección para la detección de bolitas —aunque sería un error preguntar *exactamente cuándo* ha ocurrido suficiente de esto como para que «cuente»—. No esperes que suene una campana para indicar que lo que el ojo de la rana está diciendo acaba de modificarse *ahora mismo*. No hubo un Mamífero Primario y no hay una Detección de Bolitas Primaria.

A menos que hubiera variaciones «insignificantes» o «indeterminadas» en las condiciones de los ojos de las diferentes ranas que hacen que la lengua se dispare, no podría haber materia prima (variación ciega) para la selección de un *nuevo* propósito para actuar sobre él. La indeterminación que Fodor (y otros) ve como un defecto en las explicaciones darwinianas de la evolución del significado es de hecho un requisito de cualquier evolución así. La idea de que debe de haber *algo determinado* que el ojo de la rana *realmente* quiere decir —alguna posible proposición en raniano incognoscible que exprese *exactamente* lo que el ojo de la rana le está diciendo a su cerebro— es tan sólo esencialismo aplicado al significado (o a la función). El significado, como la función, de la que tan directamente depende, no es algo que se determine al nacer. Surge no por saltación —enormes saltos en el Espacio de Diseño— o por creación especial, sino por un (típicamente gradual) cambio de circunstancias.

48. Saltar por el espacio en la biblioteca de Babel

En 1988 le mandaron a Otto Neugebauer, el gran historiador de la astronomía, la fotografía de un fragmento de papiro griego con unos cuantos números escritos en una columna. La persona que lo mandaba, un estudioso de la Antigüedad grecorromana, no tenía la menor idea del significado de este trozo de papiro y se preguntaba si Neugebauer tendría alguna pista. El erudito de 89 años volvió a analizar línea por línea las diferencias entre los números, encontró sus límites máximos y mínimos, y determinó que el papiro tenía que ser traducción de una parte de la «columna G» de una tablilla cuneiforme babilónica en la que estaba escrito un «sistema B» babilónico de efemérides lunares (una efeméride es una tabla de valores para calcular la ubicación de un cuerpo celeste en todo momento en un periodo particular). ¿Cómo *pudo* Neugebauer hacer esta deducción propia de Sherlock Holmes? Elemental: reconoció que lo que estaba escrito en griego (una secuencia de números sexagesimales —no decimales—) era parte —¡la columna G! — de un cálculo de gran exactitud de la ubicación de la Luna que los babilonios habían estado elaborando. Hay muchas maneras diferentes de calcular una efeméride, y Neugebauer sabía que cualquiera que elaborara la suya propia de manera independiente, usando su propio sistema, no habría podido inventarse exactamente los mismos números, aunque pudieran haber estado cerca. El sistema B babilónico era excelente, así que en la traducción, por agradecimiento, se había conservado el diseño, con todos sus detalles y particularidades.[33]

Neugebauer era un gran erudito, pero probablemente tú puedas realizar una hazaña de deducción parecida siguiendo sus pasos. Supón que te mandaran una fotocopia del texto que aparece en la figura 15 y te plantearan la misma pregunta: ¿qué significa? ¿De dónde pudo haber salido?

Antes de seguir leyendo, inténtalo. Probablemente puedas resolverlo aunque no sepas leer la vieja fuente tipográfica alemana *fraktur* —¡incluso si no sabes alemán!—. Fíjate otra vez, míralo de cerca. Trata de articular las palabras con gran énfasis

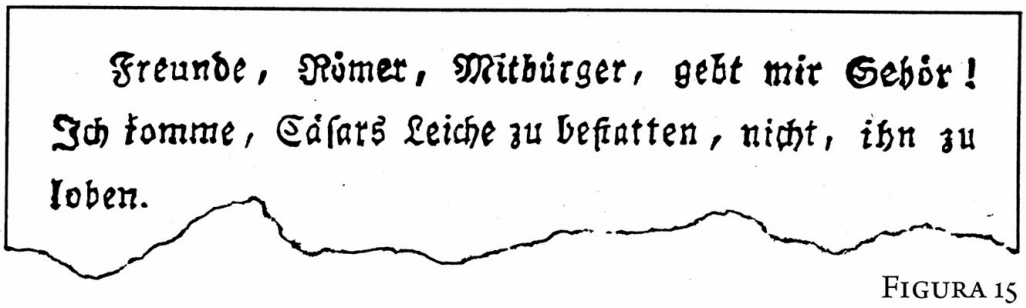


FIGURA 15

y sonoridad, y no te preocupes por la pronunciación. ¿Lo captaste? ¡Impresionante proeza! Neugebauer pudo haber tenido su columna G babilónica, pero tú rápidamente

determinaste, ¿no es así?, que este fragmento *debe de ser* parte de una traducción alemana de unas líneas de una tragedia isabelina (acto III, escena 2, líneas 79-80, para ser exactos). Ya que lo piensas, te das cuenta de que difícilmente podría ser cualquier otra cosa. Son Extremadamente bajas las probabilidades de que *esta* secuencia particular de letras alemanas se ensartara bajo ninguna otra circunstancia.[34]

La I+D que tomó crear esa secuencia es demasiado específica para que se duplique de casualidad. ¿Por qué? ¿Cuál es la particularidad que caracteriza a esa cadena de símbolos? Nicholas Humphrey enriquece la pregunta al plantear una versión más drástica: si estuvieras forzado a «relegar al olvido» una de las siguientes obras maestras, ¿cuál elegirías?: ¿los *Principia* de Newton, los *Cuentos de Canterbury* de Chaucer, *Don Giovanni* de Mozart o la Torre Eiffel? «Si estuviera obligado a hacer una elección», *responde* Humphrey:

> No tengo mayores dudas de cuál debería ser: los *Principia* tendrían que quedar fuera. ¿Y eso por qué? Porque, de todas esas obras, la de Newton es la única que no es *irremplazable*. Es muy simple: si Newton no la hubiera escrito, entonces alguien más lo habría hecho, probablemente en algunos años […] Los *Principia* fueron un glorioso monumento al intelecto humano, la Torre Eiffel fue una hazaña relativamente menor de ingeniería romántica. Sin embargo, el hecho es que mientras que Eiffel lo hizo a *su* manera, Newton simplemente lo hizo a la manera de Dios.[35]

49. ¿Quién es el autor de *Spamlet*?

Imagina que el doctor Frankenstein diseña y construye un monstruo, Spakesheare, que acto seguido se sienta a escribir una obra de teatro, *Spamlet.*

¿Quién es el autor de *Spamlet*?

Primero anotemos lo que según yo es irrelevante en esta bomba de intuición. No he dicho si Spakesheare es un robot, si está hecho de metal y chips de silicio o, como el monstruo de Frankenstein original, de tejidos humanos —o fabricado con nanotecnología a partir de células, proteínas, aminoácidos o átomos de carbono—. Con tal de que el doctor Frankenstein lleve a cabo el trabajo de diseño y construcción, para el ejemplo da lo mismo cuáles sean los materiales. Bien podría ocurrir que la única manera de construir un robot suficientemente pequeño, rápido y energéticamente eficiente para que se siente en un banco y se ponga a teclear para escribir una obra sea construirlo con células artificiales llenas de proteínas motoras de bella factura y otros nanorrobots con base de carbono. Es un interesante tema técnico y científico, pero que aquí no nos concierne. Por exactamente la misma razón, si Spakesheare es un robot de metal y silicio, puede ser más grande que una galaxia si eso se necesita para meter en su programa la complicación requerida —y puede ser que tengamos que abolir el límite de velocidad de la luz para poder imaginar que tiene lugar en una duración de dimensiones humanas—. Por lo general estas limitaciones ni se mencionan en las bombas de intuición, y por esta vez seguiremos el juego, pues nada depende de eso (gira las perillas y compruébalo). Si el doctor Frankenstein decide hacer su robot de inteligencia artificial con proteínas y cosas así, es su asunto. Si su robot puede ser fecundado por un ser humano y por lo tanto tiene posibilidades de dar a luz a una criatura e inaugurar así lo que podría decirse que es una nueva especie, es fascinante, pero lo que aquí nos incumbe es lo que suponemos la creación de Spakesheare, *Spamlet,* no sus crías. De vuelta a nuestra pregunta: ¿quién es el autor de *Spamlet*?

Para entender a fondo la cuestión tenemos que asomarnos al interior de Spakesheare y ver qué pasa ahí. En un extremo, nos encontramos ahí dentro con un archivo (si Spakesheare es un robot con una memoria de computadora) o una versión básicamente *memorizada* de *Spamlet,* completamente cargada y lista para ejecutarse. En un caso extremo como éste, el doctor Frankenstein es con toda seguridad el autor de *Spamlet* y ha usado esta creación intermedia, Spakesheare, como un simple dispositivo de almacenamiento y entrega, un procesador de palabras algo extravagante. *Todo* el trabajo de I+D se hizo antes y se copió de un modo u otro a Spakesheare.

Podemos visualizarlo más claramente si nos imaginamos a *Spamlet* y su galaxia de vecinos cercanos en la biblioteca de Babel. ¿Cómo llegó *Spamlet* ahí? ¿Cuál fue la trayectoria de I+D que la creó? Si descubrimos que el viaje se había completado para cuando se construyó y se llenó de información la memoria de Spakesheare, sabremos

que él no tuvo ningún papel en la investigación. Si nos vamos para atrás y nos damos cuenta de que la única función de Spakesheare fue hacer que un revisor ortográfico pasara por el texto almacenado antes de usarlo para dirigir sus movimientos de teclado, no nos impresionarán para nada las atribuciones de autoría de Spakesheare. Esto es una parte medible, pero evanescente, de la I+D total. Hay una galaxia, de proporciones considerables, de textos casi gemelos de *Spamlet:* aproximadamente 100 millones de mutantes menores tienen una sola errata sin corregir, y si expandimos nuestro horizonte para incluir una errata por página habremos empezado a entrar en la tierra de cantidades excesivas de variaciones sobre el tema. Si vamos un poco más atrás —cuando hemos pasado de erratas a *erramientos,*[36] esas palabras posiblemente equivocadas, o que no fueron las mejor elegidas—, hemos empezado a introducirnos en la tierra de la autoría seria, en contraste con la simple revisión de pruebas tipográficas. La relativa trivialidad de la lectura de pruebas tipográficas, y sin embargo su importancia, imposible de ignorar, en la configuración del producto final está bien representada en el Espacio de Diseño, donde cada pequeño levantamiento cuenta para algo, y a veces un pequeño levantamiento conduce a una trayectoria completamente nueva. Como de costumbre, aquí podemos citar a Ludwig Mies van der Rohe: «Dios está en los detalles».

Ahora giremos las perillas de nuestra bomba de intuición y miremos el otro extremo, en el que el doctor Frankenstein le deja la mayor parte del trabajo a Spakesheare. El escenario más realista sería con toda seguridad que el doctor Frankenstein equipara a Spakesheare con un pasado virtual, una dotación de toda una vida de seudorrecuerdos de experiencias a las que pueda recurrir cuando responda a ese obsesivo deseo de escribir una obra de teatro que le instaló Frankenstein. Podemos suponer que entre esas seudomemorias hay muchas noches en el teatro o leyendo libros, pero también algunos amores no correspondidos, algunas ocasiones en las que se salvó por un pelo, algunas traiciones vergonzosas y cosas parecidas. ¿Ahora qué pasa? A lo mejor un retazo de historia de «interés humano» en el noticiario de la tarde será el catalizador que le provoque a Spakesheare un arrebato de actividad febril en el que se pone a generar y probar: registra su memoria de arriba abajo en busca de temas útiles y transforma lo que ahí encuentra, zangolotea las piezas para crear estructuras temporales que, optimistas, compiten entre sí para ser las que lleguen a término, tras haberse visto casi todas ellas desmanteladas por los corrosivos procesos de crítica que, no obstante, de vez en cuando descubren partes provechosas, y así… Toda esta búsqueda de múltiples niveles estaría dirigida por evaluaciones de múltiples niveles generadas internamente, entre ellas la evaluación de la evaluación de las funciones de evaluación como respuesta a la evaluación de (la evaluación de) los productos de las búsquedas en curso (ciclos dentro de ciclos dentro de ciclos).

Ahora bien, si el sorprendente doctor Frankenstein realmente hubiera previsto toda

esta actividad hasta el último detalle, en el nivel más turbulento y caótico, y hubiera diseñado a mano el pasado virtual de Spakesheare y toda su maquinaria de búsqueda, para dar lugar exactamente a este producto, *Spamlet,* entonces una vez más el doctor Frankenstein sería el autor de *Spamlet,* pero también, en una palabra, Dios. Tan extrema previsión sería sencillamente milagrosa. Restableciendo una pizca de realismo a nuestra fantasía, podemos poner las perillas en una posición menos extrema y asumir que el doctor Frankenstein fue incapaz de prever todo esto en detalle, pero delegó en Spakesheare la mayor parte del trabajo duro de completar la trayectoria en el **Espacio de Diseño** hacia *una obra literaria u otra,* algo que pudiera determinar la posterior I+D que ocurriera dentro de Spakesheare mismo. Ya hemos llegado, gracias a este simple giro de la perilla, a los alrededores de la realidad misma, pues ya tenemos ejemplos reales de autores artificiales impresionantes que **Extremadamente** superan la previsión de sus propios creadores. Nadie ha creado aún una obra de teatro artificial que merezca una atención seria, pero tanto un jugador de ajedrez artificial —Deep Blue de IBM— como un compositor artificial —emi de David Cope— han alcanzado resultados que *en algunos aspectos* igualan lo mejor que el genio humano creativo puede producir.

¿Quién le ganó a Gary Kaspárov, el campeón mundial de ajedrez reinante? No Murray Campbell ni ninguno de los miembros de su equipo de IBM. Deep Blue le ganó a Kaspárov. Deep Blue diseña mejores jugadas de ajedrez que cualquiera de ellos. Ninguno de ellos puede componer una partida vencedora contra Kaspárov. Deep Blue puede. Sí, pero. Sí, pero. En este punto puedes estar tentado de insistir en que cuando Deep Blue le gana a Kaspárov en el ajedrez, sus meramente físicos métodos de búsqueda son *completamente* diferentes de los procesos exploratorios con que Kaspárov evoca sus jugadas. Sin embargo, sencillamente no es así —al menos no lo es de la única manera que podría ser significativa en el contexto de esta discusión de la perspectiva darwiniana sobre la creatividad—. El cerebro de Kaspárov está hecho de materiales orgánicos y tiene una arquitectura que en un sentido importante es distinta de la de Deep Blue, pero hasta donde sabemos sigue siendo un motor de búsqueda con una enorme capacidad para las operaciones paralelas, que con el tiempo ha elaborado un despliegue sobresaliente de técnicas heurísticas para podar ramas poco probables que sólo le harían perder el tiempo. No cabe duda de que la inversión en I+D tiene un perfil diferente en los dos casos; Kaspárov tiene métodos para extraer de partidas pasadas buenos principios de diseño, de modo que puede reconocer, y saber lo suficiente para ignorarlas, enormes porciones del espacio de juego que Deep Blue todavía tiene que escudriñar pacientemente una por una. La «percepción» de Kaspárov cambia profundamente la forma de su búsqueda, pero eso no constituye un medio de creación «*totalmente* distinto». Cada vez que las búsquedas exhaustivas de Deep Blue cierran un *tipo* de avenida que mediante algún modo algorítmico puede identificar como probablemente insignificante (una tarea difícil pero no

imposible), puede reutilizar esa I+D cuando resulte apropiado, tal como Kaspárov. Los diseñadores de Deep Blue han hecho gran parte de este trabajo analítico y se lo han dado a Deep Blue como un atributo innato, pero también Kaspárov se ha beneficiado con los frutos de cientos de miles de años persona de exploración ajedrecística que jugadores, entrenadores y libros le han transmitido, y que posteriormente se instalaron en los hábitos de su cerebro.

A este respecto es interesante contemplar una sugerencia que Bobby Fischer hizo una vez. Propuso que, para que el juego del ajedrez recuperara su pretendida pureza racional, al principio de cada partida las principales piezas blancas se colocaran *aleatoriamente* en la última fila y las principales piezas negras se acomodaran en el mismo orden aleatorio (en reflejo) en la fila opuesta (pero siempre asegurándose de que cada lado tuviera un alfil en escaque blanco y uno en escaque negro, y que el rey estuviera entre las torres). Esto de inmediato volvería casi totalmente obsoleta la montaña de aperturas memorizadas, tanto para las personas como para las máquinas, pues muy rara vez entraría en juego esta sabiduría popular. Uno se vería obligado a volver a confiar en los principios fundamentales; tendría que hacer más trabajo duro de diseño en el momento, con el reloj caminando. Está lejos de ser claro si este cambio en las reglas beneficiaría más a los seres humanos o a las computadoras. Todo depende de qué clase de jugador de ajedrez se base en mayor medida en lo que, en efecto, es pura memorización: dependencia *con mínima comprensión* de la I+D de exploradores anteriores.

El hecho es que el espacio de búsqueda en el ajedrez es demasiado grande para explorarse exhaustivamente en el momento, hasta para Deep Blue, así que, como Kaspárov, poda sus árboles de búsqueda corriendo riesgos calculados, y, como Kaspárov, a menudo los tiene calculados de antemano. Cabe suponer que ambos hacen inmensas cantidades de cómputos meramente físicos en sus muy diferentes arquitecturas. A fin de cuentas, ¿qué saben de ajedrez las neuronas? Cualquier trabajo que *ellas* hagan debe ser trabajo meramente físico de una u otra clase.

Puede parecer que estoy dando por sentado un enfoque informático y de inteligencia artificial al describir así el trabajo que hace el cerebro de Kaspárov, pero el trabajo tiene que hacerse de algún modo, y jamás se ha expresado *otra* manera de hacerlo. De nada serviría decir que Kaspárov usa su «perspicacia» o «intuición», pues eso sólo significaría que Kaspárov mismo no tiene acceso privilegiado (perspicacia) al modo como le llegan los buenos resultados. Así, como casi nadie sabe cómo lo hace el cerebro de Kaspárov —mucho menos Kaspárov mismo—, todavía no hay ningún elemento de prueba que apoye la afirmación de que los medios de Kaspárov son «totalmente distintos» de los que aprovecha Deep Blue. Uno debería recordarlo cuando se sintiera tentado de insistir en que «por supuesto» los métodos de Kaspárov son enormemente distintos. ¿Qué demonios podría mover a alguien a hacer esa clase de afirmaciones arriesgadas?

¿Pensamiento desiderativo? ¿Miedo?

Pero eso sólo es ajedrez, dirás, no arte. El ajedrez es *trivial* comparado con el arte (ahora que el campeón mundial de ajedrez es una computadora). Es aquí donde los experimentos sobre Inteligencia Musical, o EMI por sus siglas en inglés, del compositor y *hacker* David Cope, entran en juego.[37] Cope se propuso crear un simple aparato para mejorar la eficiencia, un auxilio para compositores que ayudara a superar los bloqueos a los que se enfrenta cualquier creador, una extensión de alta tecnología de los vehículos de búsqueda tradicionales (el piano, el papel pautado, la grabadora, etc.). A medida que EMI se hizo más competente, se ascendió a sí mismo a compositor hecho y derecho, que incorporaba cada vez más procesos de generar y probar. Cuando se alimenta a EMI con música de Bach, responde generando composiciones musicales en el estilo de Bach. Cuando se le da Mozart, Schubert, Puccini o Scott Joplin, inmediatamente analiza sus estilos y luego compone nueva música en ellos, mejores pastiches que los que Cope mismo —o que prácticamente cualquier compositor humano— pueda escribir. Cuando se le alimenta de música de dos compositores, EMI puede rápidamente componer piezas inquietantes que unen sus estilos, y cuando se le dan al mismo tiempo todos estos estilos (sin haber limpiado el paladar, podría decirse), procede a escribir música basado en la totalidad de su «experiencia» musical. Luego pueden usarse las composiciones resultantes para alimentarlo de vuelta, una y otra vez, junto con cualquier otra música que se encuentre en formato MIDI,[38] y el resultado es el propio estilo musical «personal» de EMI, uno que reconoce abiertamente su deuda con los grandes maestros sin dejar de ser una integración sin lugar a dudas idiosincrásica de toda esta «experiencia». EMI ahora puede componer no sólo invenciones de dos partes y arias sino sinfonías enteras —y lo último que supe fue que ya ha compuesto más de mil—. Son lo bastante buenas para engañar a los expertos (compositores y profesores de música), y yo mismo puedo dar fe de que con un aria de EMI-Puccini se me hizo un nudo en la garganta. David Cope no puede proclamarse compositor de sinfonías, motetes y arias, así como Murray Campbell no puede decir que ha vencido a Kaspárov en el ajedrez.

Para un darwiniano, este nuevo elemento en la cascada de grúas no es sino lo más reciente en una larga historia, y deberíamos reconocer que el límite entre los autores y sus artefactos debería ser tan penetrable como todos los otros límites en la cascada. Cuando Richard Dawkins observa que el dique del castor forma parte del fenotipo del castor —su *fenotipo extendido*— tanto como sus dientes y su pelaje, monta el escenario para la observación de que los límites de un autor humano se pueden extender exactamente del mismo modo.[39] De hecho, esto lo sabemos desde hace siglos y hemos creado diversas convenciones semiestables para lidiar con los productos de Rubens, o el *taller* de Rubens, o los varios discípulos de Rubens. Si en una situación puede echarse una mano, siempre podemos plantear la pregunta de exactamente quién está ayudando a

quién, qué es creador y qué es creación.

50. Ruido en el hotel virtual

Piensa en la diferencia entre mundos virtuales y mundos reales. Si quieres hacer un hotel real, tienes que dedicar tiempo, energía y materiales para conseguir que la gente no pueda oír qué pasa en la habitación de junto. Cuando haces un hotel virtual obtienes gratuitamente ese aislamiento sonoro. En un hotel virtual, si quieres que la gente pueda oír al de la habitación de junto tienes que agregar esa capacidad. Tienes que agregar *no* aislamiento. También tienes que agregrar sombras, olores, vibración, polvo, huellas digitales y desgaste natural. Todas estas características no funcionales son gratuitas en el mundo real y concreto. El término genérico de lo que debe agregarse a los mundos virtuales para hacerlos más realistas es *detección de colisión.* Si alguna vez has tratado de hacer un videojuego para computadora, probablemente pronto te diste cuenta de que no basta con hacer que las figuras de la pantalla se muevan. Las figuras se atraviesan unas a otras sin efecto alguno a menos que les incorpores la detección de colisión en el bucle de actualización (un ciclo de los programas que incesantemente interviene en lo que el objeto esté haciendo y pregunta «*¿No estoy colisionando con nada?»*).

En su libro *Le Ton beau de Marot,* Doug Hofstadter hace notar el papel de lo que llama *intromisiones espontáneas* en los procesos creativos.[40] En el mundo real, casi cualquier cosa que ocurra deja una estela, hace sombra, huele a algo, hace ruido, y esto proporciona oportunidades en abundancia para las intromisiones espontáneas. También es precisamente lo que escasea en el mundo virtual. De hecho, uno de los encantos de los mundos virtuales desde el punto de vista de los modeladores informáticos es ese silencio: nada pasa excepto lo que tú provees, de uno u otro modo. Esto te permite empezar de cero y agregar una por una las características a tu modelo, y ver cuál es el modelo mínimo que producirá los efectos buscados.

Esta ausencia de ruido vuelve sumamente limitadas las simulaciones de la evolución por computadora, pues la evolución por selección natural se alimenta de ruido: convierte el ruido fortuito con el que se topa en señal, chunches en herramientas, errores de programación en características. Una de estas primeras simulaciones de la evolución, y que sigue siendo una de las más impresionantes, es *Evolved Virtual Creatures* de Karl Sims, que puede verse en <karlsims.com/evolved-virtual-creatures.html>.

Sims empezó con *ensamblajes aleatorios* de bloques articulados virtuales con músculos virtuales para mover las articulaciones, y luego dejó que evolucionaran en su mundo virtual con su física virtual. El programa automáticamente seleccionaba los ensamblajes que más lejos llegaban en sus ensayos de apareamiento virtual, y luego repetía el ciclo con la descendencia resultante. Evolucionaron nadadores, caminadores y saltadores cada vez mejores, sin ninguna intervención del Diseñador Inteligente. Los diseños, nada aleatorios, resultantes demostraron cuán eficientemente la evolución (virtual) podría **maomeno** descubrir buenos principios de diseño y **maomeno** reinventar

una asombrosa variedad de características que se encuentran en la naturaleza.

Es un ejemplo espectacular de todo lo que puedes obtener de un modelo *relativamente* simple, pero también muestra los estrechos límites de la evolución en el mundo virtual. Sims diseñó el simple sistema «de desarrollo» que toma genomas enteros como entradas y arroja organismos originales como salidas, pero todo ese proceso ocurre tras bambalinas y no forma parte del mundo virtual que se está simulando. En consecuencia, un golpe, choque o colisión accidental con un trozo de escombros (virtuales) no puede acortar o alargar un genoma, o cambiar las reglas de expresión del gen. Toda esa maquinaria simplemente no está en el mundo virtual con sus cosas virtuales, y por lo tanto es inmutable. Por ejemplo, no hay manera de que una criatura de Sims produzca por evolución un nuevo cromosoma. Todo su sistema genético está fuera del modelo: no enfrenta la selección natural misma, sino que simplemente lleva información genética por decreto entre generaciones (para otro ejemplo de este fenómeno, véase el apartado 51).

En un modelo informático de creatividad debería haber tirados por ahí chunches con los que tus procesos creativos pudieran tropezarse, ruidos que tus procesos creativos no pudieran evitar oír. La intrusión espontánea de ese ruidito de la habitación de junto puede alterar lo que esos procesos están haciendo, de manera inesperada (por serendipia) o de manera destructiva, pero sea como sea esto abre nuevas posibilidades. El aprovechamiento de los accidentes es la clave de la creatividad, ya sea que se esté haciendo un genoma, una nueva conducta o una nueva melodía.

Permíteme aclarar lo que *no* estoy diciendo. El problema con las criaturas evolucionadas de Sims no es que no estén hechas de carbono o que no contengan proteínas o hemoglobina. El problema es que son virtuales. Y por ser virtuales, viven en un mundo que es muchos órdenes de magnitud más simple que el mundo de la evolución biológica. Lo mismo puede decirse de EMI de Cope. Maravilloso como es, EMI es órdenes de magnitud más simple que el mundo de la composición musical humana. Lo bonito de ambos casos es el descubrimiento de cuánto puedes obtener de algo tan limpio, tan libre de ruido, tan abstracto.

Podemos imaginar que para mejorar EMI de Cope, el trabajo de Sims o cualquier otro proyecto por el estilo sobre vida artificial o creatividad artificial, agregamos en ese mundo más y más y más chunches, más y más oportunidades para las colisiones. Esto proporcionará más cosas virtuales para que sus simulaciones se las arreglen con ellas, y nunca se sabe cuándo puede llegar la serendipia, si se le da la oportunidad. Sin embargo, piensa cómo podría verse ese consejo antiintuitivo: «Modeles lo que modeles, asegúrate de que todo fenómeno, toda subrutina, todo lo que pasa en ese mundo difunda una variedad de efectos no funcionales a través del mundo: que haga ruidos superfluos, deje una estela, levante polvo, cause vibraciones, etc.» ¿Por qué? ¿Para qué es todo este

ruido? No es para nada: simplemente está ahí, de modo que cualquier otro proceso tenga ese ruido como fuente potencial de señal, como algo que, por la alquimia del algoritmo creativo, *podría* convertirse en función, en arte, en significado. Cada incremento de diseño en el universo empieza con un momento de serendipia, la intersección no diseñada de dos trayectorias que producen algo que retrospectivamente resulta ser más que una simple colisión. Sin embargo, en la medida en que los modeladores informáticos siguen este consejo, desaprovechan la eficiencia que hace de las computadoras tan grandes herramientas. Así, tenemos aquí una especie de homeostasis. Puede verse que, no por ninguna razón misteriosa, el modelado informático de la creatividad se enfrenta a rendimientos decrecientes. Para acercarse cada vez más a la creatividad de un compositor humano, el modelo tiene que volverse cada vez más concreto; tiene que modelar más y más de las colisiones incidentales que inciden en un compositor encarnado.

51. Herb, Alice y Hal, el bebé

El difunto gran teórico evolucionista George Williams insistía en que era un error identificar genes con moléculas de ADN. Sería más o menos equivalente a pensar que *Hamlet* está hecho de tinta. Por supuesto que cualquier ejemplar real de la obra de Shakespeare tiene que estar hecho de algo (si no de tinta, entonces tal vez dibujos con forma de letra en la pantalla de una computadora, o hasta cadenas de código binario quemadas en un disco compacto), pero la obra es una cosa abstracta e informativa que puede brincar de un medio a otro. Los genes, como recetas para hacer proteínas, según esta manera de pensar —que siempre me ha dado la impresión de ser la correcta— también son cosas abstractas e informativas. Sin embargo, ha habido quienes no están de acuerdo, quienes dudan del valor de concebir así a los genes. Para ellos, y en particular para el filósofo de la biología Peter Godfrey-Smith, he elaborado una pequeña bomba de intuición:

Herb y Alice quieren tener un bebé, y así es como lo hacen:

1. Los dos encargan que se secuencien sus genomas. Reciben por correo un archivo de datos que consiste en sus genomas deletreados en dos secuencias de aproximadamente 3 000 millones de letras A, C, G, T…
2. Luego escriben un pequeño programa informático que hace el algoritmo de la meiosis en sus respectivos genomas (aleatoriamente), con lo que se crean espermatozoides y óvulos virtuales, que luego se unen (aleatoriamente) *in silico* para crear una nueva especificación genómica (que pasa todas las pruebas forenses para asegurarse de que es la especificación del ADN de un vástago de Herb y Alice). (Hasta este punto todo se ha hecho literalmente de acuerdo con los símbolos A, C, G, T, un proceso meramente informático de reescritura de cadenas.)
3. Luego esta especificación se usa para construir, codón por codón, una implementación de ADN real de todo el genoma: A = adenina, C = citosina, G = guanina y T = timina (el laboratorio de Craig Venter ya puede hacerlo).
4. Entonces este genoma se implanta en el núcleo de un óvulo humano (¿importa de quién sea el óvulo, si su propio ADN se quitará antes de la implantación en el núcleo?), y se convierte en un «bebé de probeta» en una de las formas habituales.

El bebé resultante, Hal, ¿es hijo de Herb y Alice? A mí me parece de lo más evidente que Hal es en efecto su descendencia biológica, pues aprovecha toda la información genética que aportarían si la concepción hubiera sido de la manera normal. Esta bomba de intuición destaca lo que importa en reproducción: la información y la

transmisión causal de la información (en este caso, la forma del código ASCII para «A», «C», «G» y «T», no en forma de moléculas). El enlace causal podría, por ejemplo, pasar a través de satélites de comunicación, en vez de tomar las rutas bioquímicas más directas.[41]

Godfrey-Smith coincidía conmigo en que Hal es descendiente de Herb y Alice, pero tenía algunas reservas sobre mi manera de plantearlo.[42] En el mismo espíritu de crítica constructiva, yo reconocí que *hay* una diferencia biológica importante entre el proceso que Herb y Alice usaron para procrear y el que normalmente usamos los demás. ¿Y si todo el mundo lo hiciera a la manera de Herb y Alice? Como la información genética de Herb no se transporta en su vehículo habitual —un espermatozoide— para llegar al óvulo, la *movilidad de los espermatozoides* ya no estaría sometida a la presión selectiva y, si no intervienen otros factores, disminuiría con las generaciones. Úsala o piérdela. De todas formas sostengo —y creo que esta bomba de intuición lo ilustra claramente— que la estructura molecular que permanece constante en líneas generales a través de las generaciones se conserva *porque* encarna la información que encarna.

Si hubiera estructuras alternativas que conservaran la información, la evolución continuaría libre de trabas. Este aserto puede examinarse más a fondo en otra bomba de intuición. Imagina que en otro planeta las generaciones «de número non» usaran los peldaños de ADN A, C, G y T, y las generaciones «de números pares» usaran una doble hélice distinta, llamémosla XNA, con peldaños hechos de P, Q, R y S (algunas otras moléculas). Podemos suponer que las moléculas XNA de la descendencia estuvieran hechas de plantillas de ADN de sus padres mediante un mecanismo parecido al ARN [ácido ribonucleico] mensajero, pero que «tradujera» de uno al otro los dos diferentes lenguajes bioquímicos. Otro mecanismo mensajero en la siguiente generación traduciría de regreso los lenguajes, etc. Si quisieras descendencia, tendrías que encontrar una pareja cuyos genes estuvieran escritos en el mismo lenguaje, pero esa descendencia tendría genomas escritos en el otro lenguaje. Las uniones edípicas serían infértiles —lo cual probablemente está bien—, pero podría haber muchas tragedias tipo Romeo y Julieta en las que los amantes no podrían procrear por provenir de comunidades opuestas (siempre podrían acomodarse a una vida sexual estéril y adoptar bebés de cualquiera de los tipos, o incluso hacer de donadora de óvulos y donador de esperma y criar medios hermanos en abundancia). En un mundo así, aparte de estas complicaciones, la evolución proseguiría como siempre, transmitiendo información genética sobre adaptaciones valiosas (y enfermedades hereditarias, etc.), a través de las generaciones, en sistemas de codificación alternativos que podrían ser tan estructuralmente distintos como quisieras. Los mismos genes, diferentes moléculas. Cada gen tendría dos formas, tan diferentes como «gato» y «*chat»,* «casa» y «*maison»* (observa el paralelismo con el ejemplo de las

dos cajas negras: en ambos casos lo que los dos vehículos sintáctica y estructuralmente distintos tienen en común es la *misma información,* la misma semántica).

52. Memes

Todavía no he mencionado a los memes, lo que a algunos lectores puede haberles hecho preguntarse si los abandoné. Para nada. El concepto de *meme* es una de mis herramientas de pensamiento favoritas, y tengo mucho que decir sobre él —¡demasiado para que quepa en este libro!—. Ya he dicho mucho acerca de los memes en otros lugares.[43] Por diversas razones, mucha gente tiene un disgusto visceral por el concepto, así que tienden a tragarse las numerosas críticas que se han planteado. He decidido que debería tratar de defender a los memes una vez más y lidiar con todos sus críticos, tanto los serios como los que, blancos de espanto, simplemente odian a los memes, pero a eso le dedicaré un librito entero. Mientras tanto, quienes quieran saber más sobre memes échenle un ojo a mi ensayo sobre el tema, «The New Replicators».[44]

Sin embargo, como un adelanto, presento aquí una breve introducción al concepto *serio* de memes (en contraste con el empleo laxo que han sobrepopularizado los moradores de internet). Como Dawkins señaló cuando introdujo el concepto de meme como artículo cultural que se copia, el fundamento principal de la biología es

> que toda la vida evoluciona mediante la supervivencia diferencial de entidades reproductoras […]
>
> El gen, la molécula de ADN, resulta ser la entidad reproductora prevaleciente en nuestro planeta. Puede haber otras. Si las hay, siempre y cuando también se cumplan otras condiciones, casi siempre tenderán a convertirse en la base de un proceso evolutivo.
>
> ¿Pero tenemos que ir a mundos distantes para encontrar otras clases de reproducción y otras consecuentes clases de evolución? Creo que en este mismo planeta recientemente ha surgido una nueva clase de reproductor. Lo tenemos delante de las narices. Aún está en su primera infancia, sigue dejándose llevar torpemente por la corriente en su caldo primigenio, pero ya está consiguiendo cambio evolutivo a una velocidad que deja al viejo gen jadeando muy atrás.[45]

Dos intuiciones principales que surgen de esta herramienta de pensamiento alteran drásticamente el paisaje de nuestra imaginación cuando pensamos en la cultura y la creatividad humanas. En primer lugar, los memes hacen añicos la por otra parte seductora idea de que sólo hay dos caminos al buen diseño: o los *genes* o el *genio*. Para la mayoría de los pensadores, hasta que los memes les abran los ojos, si algo en la vida humana presenta los signos que revelan adaptación de los medios a los fines o eficiencia funcional, debe de ser producto de la selección natural *genética* o bien de un

pensamiento humano deliberado, amplio, propositivo, exhaustivo, con propósitos: diseño *inteligente*. Podría parecer que la Segunda Ley de Orgel —la evolución es más lista que tú— encierra las dos opciones, pero de hecho hay una tercera posibilidad que en todas partes encontramos ejemplificada: la selección cultural, no genética, que se consigue mediante el mismo proceso de selección natural mecánica que nos da los genes. Una observación de los pescadores bretones en la isla de Groix que tiene más de cien años nos da un vívido ejemplo: «Todos los botes se copian de otro bote […] Es el mismo mar el que moldea los botes, eligiendo los que funcionan y destruyendo los demás».[46] Tan claro como el día, esto es selección natural; los isleños tienen una sencilla regla: si regresa intacto del mar, ¡cópialo! Podrán tener una notable comprensión de los principios de arquitectura naval que retrospectivamente respaldan sus diseños favoritos, pero es estrictamente innecesaria. La evolución se encarga del control de calidad. Lo mismo se aplica a las reglas gramaticales, las palabras, las prácticas religiosas y muchas otras características fundamentales de la cultura humana: *nadie* las diseñó, y no «están en nuestros genes», y con todo, se diseñaron excelentemente.

La segunda intuición es que el precio que pagamos por tener esta autopista extra de la información, este copioso medio de diseño y transmisión del que ninguna otra especie goza, es que los memes *tienen su propia aptitud,* igual que todos los demás simbiontes que prosperan en nuestra compañía, y su aptitud es hasta cierto punto independiente de la nuestra. La ceguera a esta idea es endémica, y es particularmente evidente cuando la gente discute las explicaciones evolutivas de la religión: «Oh, así que estás trabajando en una teoría evolutiva de la religión. ¿Qué beneficio crees *tú* que acarrean las religiones? *Deben* de ser buenas para algo, si al parecer todas las culturas humanas tienen religión de alguna u otra forma». Pues bien, toda cultura humana tiene también el resfriado común. ¿Qué beneficio acarrea? Es bueno para sí mismo. Deberíamos prepararnos para encontrar reproductores culturales que no son benéficos pero que sin embargo se las arreglan para prosperar. Esto nivela el campo de juego de las teorías de la evolución cultural, al remplazar la idea estrecha de que las innovaciones culturales —igual que las innovaciones genéticas— siempre mejoran la aptitud de quienes las transmiten. Los memes son simbiontes informativos, y como los billones de simbiontes mutualistas que también nos habitan, no podríamos vivir sin ellos, pero eso no significa que todos sean nuestros amigos. Algunos son plagas dañinas que no nos harían ninguna falta.

Resumen

En este capítulo he tratado de demostrar que el pensamiento darwiniano sí está a la altura de lo que se espera de él como el ácido universal: voltea de cabeza el mundo tradicional en su totalidad al poner a prueba la imagen jerárquica de diseños que surgen de ese genio de genios, el Diseñador Inteligente, y remplazarlo con la emocionante imagen de procesos cíclicos mecánicos, carentes de motivos, que producen grandes cantidades de combinaciones cada vez más resistentes hasta que empiezan a reproducirse por sí mismos y aceleran el proceso de diseño reutilizando los mejores pedazos una y otra vez. Algunos de estos primeros descendientes a la larga unen fuerzas (una grúa de fundamental importancia, la simbiosis), lo que lleva a la multicelularidad (otra grúa importantísima), la cual lleva a vehículos de exploración más efectivos que son posibles por la reproducción sexual (grúa muy importante también) y tarde o temprano conducen al lenguaje y a la evolución cultural (igualmente grúas) en una especie, que proporcionan el medio para la literatura, la ciencia y la ingeniería, las últimas grúas que han surgido, y que a su vez nos permiten «ponernos *meta-*» de una manera que ninguna otra forma de vida puede, y reflexionar de muchas maneras sobre quiénes somos y qué somos, y cómo llegamos aquí, modelando estos procesos en obras de teatro y novelas, teorías y simulaciones por computadora, y cada vez más herramientas de pensamiento para agregar a nuestra impresionante caja de herramientas.

Esta perspectiva es tan ampliamente unificadora y al mismo tiempo tan generosa con las intuiciones detalladas que podría decirse que es por sí sola una herramienta eléctrica. A los que extrañamente aún les repele el pensamiento darwiniano deben considerar la posibilidad de que si tratan de hacerlo por su cuenta con tan sólo las herramientas de mano de la tradición, se encuentren trabajando lejos de la vanguardia de la investigación en importantes fenómenos tan diversos como epidemiología y epistemología, biocombustibles y arquitectura del cerebro, genética molecular, música y moralidad.

[1] Algunos comentaristas han pensado que mi comparación de la idea de Darwin con el ácido universal se inspiró en el ADN, que a fin de cuentas es ácido *d*esoxirribonucleico, pero me refiero a algo más universal: el ADN no es el único medio de la evolución en nuestro planeta, y quién sabe qué otros ejemplos de evolución habrá en todo el universo.

[2] Jorge Luis Borges, *Ficciones,* Alianza, Madrid, 1997.

[3] Borges eligió números un poco diferentes: los libros tenían 410 páginas, con 40 renglones de 80 caracteres. El número total de caracteres por libro se acerca tanto al mío (1 312 000 contra 1 000 000) que da lo mismo. Yo elegí números redondos porque son más fáciles de manejar. Borges eligió un conjunto de caracteres de tan sólo 25 elementos, que es suficiente para un español de puras mayúsculas (con espacio, coma y punto como toda puntuación), pero no para el inglés. Escogí el más espacioso 100 para que no quepa duda de que hay lugar para las letras mayúsculas y minúsculas y para la puntuación de todas las lenguas que usan el abecedario latino.

[4] S. Hawking, en *A Brief History of Time* (Bantam, Nueva York, 1988, p. 129), insiste en expresarlo así: «Hay aproximadamente diez millones de millones de millones de millones de millones de millones de millones de millones de millones de millones de millones de millones de millones (un 1 con 80 ceros) de partículas en la región del universo que podemos observar». M. Denton (*Evolution: A Theory in Crisis,* Burnett Books, Londres, 1985) proporciona el cálculo aproximado de 10^{70} átomos en el universo observable. Manfred Eigen, en *Steps Towards Life* (Oxford University Press, Oxford, 1992, p. 10), calcula el volumen del universo en 10^{84} centímetros cúbicos.

[5] La biblioteca de Babel es finita pero, aunque parezca mentira, contiene entre sus paredes todas las oraciones gramaticales del español. ¡Pero eso es un conjunto infinito, mientras que la biblioteca es finita! De todas formas, cualquier oración del español, de la extensión que sea, puede descomponerse en trozos de 500 páginas, cada uno de los cuales está en algún lugar de la biblioteca. ¿Cómo es posible? Algunos libros pueden utilizarse más de una vez. El ejemplo más despilfarrador es el más fácil de entender: como hay volúmenes que tienen un solo carácter y todo el resto está en blanco, el uso repetido de estos cien volúmenes creará cualquier texto de cualquier extensión. Como Quine señaló en su informativo y entretenido ensayo «Universal Library», si aprovechas esta estrategia de reutilizar volúmenes, y traduces todo al código ASCII que use tu procesador de palabras, ¡puedes almacenar toda la biblioteca de Babel en dos volúmenes sumamente delgados, en uno de los cuales está impreso un 0 y en el otro de los cuales aparece un 1! (Quine también señala que Gustav Theodor Fechner, el psicólogo, presentó la fantasía de la biblioteca universal mucho antes que Borges).Véase W. V. O. Quine, «Universal Library», en *Quiddities: An Intermittently Philosophical Dictionary,* Harvard University Press, Cambridge, 1987.

[6] Quine acuña el término «hiperastronómico» para el mismo propósito.

[7] P. De Vries, *The Value of Laughter,* Little Brown, Boston, 1953. El texto continúa: «Con toda libertad. Llámeme a cualquier hora del día o de la noche…» De Vries también pudo haber inventado el juego de ver qué tan grande efecto (perjudicial o no) puedes conseguir con un solo cambio tipográfico. Algunos de los mejores: «Whose woods are these, I think I know; his house is in the *V*illage though» [primeros versos del poema de Robert Frost «Stopping by Woods on a Snowy Evening»: «De quién son estos bosques, creo saberlo; su casa está en la aldea, empero». Al poner mayúscula la «V» de *village,* la casa se traslada de la aldea a The Village, como se conoce a Greenwich Village, barrio residencial de Manhattan (T.)]. Otros han seguido el juego. En el estado de naturaleza, nos dice el Hobbes mutante, «la esposa del hombre es solitaria, pobre, tosca, embrutecida y breve» [«the *w*ife of man, solitary, poore, nasty, brutish, and short», resultado de cambiar la «l» de «life», vida, como aparece originalmente en el *Leviathan,* por «w» y obtener «wife», esposa]. O piensa en la pregunta «Am I my brothel's keeper?» [«brother», «hermano», se convierte en «brothel», «burdel», en las palabras de Caín en el Génesis 4:9: «¿Acaso soy el guardián de mi hermano?» (T.)].

[8] R. Dawkins, *The Blind Watchmaker,* Longmans, Londres, 1986.

[9] M. Ridley, *The Red Queen: Sex and the Evolution of Human Nature,* Macmillan, Nueva York, 1993, p. 45.

[10] R. Dawkins, *The Ancestor's Tale: A Pilgrimage to the Dawn of Time,* Weidenfeld & Nicolson, Londres,

2004, pp. 155-156. © 2004 Richard Dawkins. Usado con permiso de Houghton Mifflin Harcourt Publishing Company y The Orion Publishing Group. Todos los derechos reservados.

11 R. B. MacKenzie, *The Darwinian Theory of the Transmutation of Species Examined,* Nisbet, Londres, 1868.

12 R. Dawkins, *Climbing Mount Improbable,* Viking Penguin, Londres, 1996, p. 457.

13 R. Dawkins, *The Ancestor's Tale, op. cit.,* p. 457.

14 C. Schönborn, «Finding Design in Nature», *The New York Times,* Editorials/op-Ed, Nueva York, 7 de julio de 2005.

15 A. Ruina, «Cornell Ranger, 2011, 4-Legged Bipedal Robot 40.5 Mile Ultra-Marathon Pictures», Cornell University, Ithaca, 2011.

16 Bueno, no del todo imposibles. Los satélites geoestacionarios, que orbitan alrededor de la Tierra al unísono con su rotación, son una especie de gancho celeste real y no milagroso. Lo que los hace tan valiosos —lo que los convierte en inversiones financieramente sólidas— es que a menudo sí tenemos muchas ganas de colgar algo muy alto, en algún lugar del cielo (algo como una antena, una cámara o un telescopio). Los satélites no son prácticos para *levantar,* ¡ay!, porque tienen que colocarse tan alto en el cielo. La idea se ha analizado cuidadosamente. Resulta que una cuerda de la fibra artificial más resistente que se haya inventado tendría que tener más de 100 metros de diámetro hasta arriba —en su descenso podría estrecharse hasta convertirse en un sedal casi invisible— sólo para aguantar su propio peso, deja tú cualquier carga. Aunque pudiera hilarse un cable así, ¡no querríamos que se saliera de órbita y cayera en alguna ciudad!

17 Me dio mucho gusto cuando se inventó la *grúa celeste* para ayudar a bajar a la superficie de Marte a Curiosity, el explorador robótico, el 6 de agosto de 2012, y me dio aún más gusto que la llamaran *grúa* celeste, no *gancho* celeste, ya que no es producto de ningún milagro sino de una ingeniería deslumbrante.

18 J. Holland, *Adaptation in Natural and Artificial Systems,* University of Michigan Press, Ann Arbor, 1975.

19 J. M. Smith, *The Evolution of Sex,* Cambridge University Press, Cambridge, 1978.

20 S. J. Gould, *Ever since Darwin,* W. W. Norton & Company, Nueva York, 1977, p. 99.

21 A. Zahavi, «The Theory of Signal Selection and Some of Its Implications», en V. P. Delfino (comp.), *Proceedings of the International Symposium on Biology and Evolution,* Adriatica Editrice, Bari, 1987, pp. 305-327, y C. D. FitzGibbon y J. H. Fanshawe, «Stotting in Thomson's Gazelles: An Honest Signal of Condition», *Behavioral Ecology and Sociobiology,* 23 (2): 69-74, 1988.

22 Tomado, con modificaciones, de D. H. Sanford, «Infinity and Vagueness», *Philosophical Review,* 84 (4): 520-535, 1975.

23 R. Dawkins, *The Ancestor's Tale, op. cit.*

24 R. L. Cann, M. Stoneking y A. C. Wilson, «Mitochondrial dna and Human Evolution», *Nature,* 325: 31-36, 1987.

25 Observa una importante diferencia entre los legados de la Eva Mitocondrial y el Adán Y-cromosómico: todos, hombres y mujeres, tenemos mitocondrias en nuestras células, y todas vienen de nuestras madres; si eres hombre, tienes un cromosoma Y y lo obtuviste de tu padre, mientras que prácticamente ninguna mujer —aunque algunas sí— tiene un cromosoma Y.

26 Dejando de lado la posibilidad lógica y fijándonos en la historia real, podemos ver que ha habido hombres que trataron de monopolizar el engendramiento en sus poblaciones, y por lo visto en ocasiones lo consiguieron en un grado sorprendente: reyes y guerreros poderosos que poseyeron y preñaron a cientos de mujeres. Para bien o para mal, probablemente hicieron una desproporcionada contribución a nuestro acervo genético.

27 Dale Peterson me ha hecho notar el hecho de que los martillos de piedra de los chimpancés se pulen gradualmente por un uso continuo a lo largo de varios periodos de tiempo, y esto ofrece un buen «eslabón perdido» en una serie que comienza con levantar una piedra más o menos al azar y golpear una nuez con ella, y luego escoger la piedra con la mejor forma para ese trabajo que puedas encontrar en los alrededores, y luego

buscar más lejos una piedra con mejor forma, y luego darte cuenta (a simple vista y apoyado en la memoria visual) de que tu piedra favorita está mejorando gracias al uso, hasta que finalmente alcanzas la sofisticación de disponerte a darle forma a la piedra para una tarea que tienes en mente.

28 D. Bray, *Wetware: A Computer in Every Living Cell,* Yale University Press, New Haven, p. 75.

29 Para un brillante estudio, véase A. Clark, «Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science», *Behavioral and Brain Sciences,* 36 (3): 181-204, 2013.

30 M. J. Behe, *Darwin's Black Box: The Biochemical Challenge to Evolution,* Free Press, Nueva York, 1996.

31 El papel de los ciclos abióticos en cambiar la probabilidad de que surjan células reproductivas se discute con más detalle en mi «Evolution of Reasons», en B. Bashour y H. D. Muller (comps.), *Contemporary Naturalism and Its Implications, op. cit.*

32 Jerome Y. Lettvin, U. Maturana, W. McCulloch y W. Pitts, «What the Frog's Eye Tells the Frog's Brain», *Proceedings of the Institute of Radio Engineers,* 47 (11): 1940-1951, 1959.

33 Véase O. Neugebauer, «A Babylonian Lunar Ephemeris from Roman Egypt», en E. Leichty, M. de J. Ellis y P. Gerardi (comps.), *A Scientific Humanist: Studies in Honor of Abraham Sachs,* Occasional Publications of the Samuel Noah Kramer Fund núm. 9, Filadelfia, 1989, pp. 301-304.

34 Aquí está el fragmento de alemán en una fuente moderna más fácil de leer: *Freunde, Römer, Mitbürger, gebt mir Gehör! Ich komme, Cäsars Leiche zu bestatten, nicht, ihn zu loben* («Amigos, romanos, conciudadanos, ¡escúchenme! Vengo a enterrar el cadáver de César, no a alabarlo» [W. Shakespeare, *Julio César*]).

35 N. Humphrey, «Scientific Shakespeare», *The Guardian,* Londres, 26 agosto de 1987.

36 [En inglés, de *typos* a *thinkos;* sobre este último término dice Dennett:] Esta espléndida acuñación es la manera como un *hacker* de computadora se refiere a un error de codificación en el nivel semántico, no el sintáctico. Dejar fuera un paréntesis es una errata *[typo];* olvidarse de declarar una variable local es un *erramiento [thinko]*. En cualquier actividad humana donde haya una interpretación semántica o intencional, y claros cánones de corrección o elegancia, hay lugar para los *erramientos.* Describir a una dama como meretricidora en vez de merecedora es un *erramiento,* no una errata. Véase D. C. Dennett, «From Typo to Thinko: When Evolution Graduated to Semantic Norms», en S. C. Levinson y P. Jaisson (comps.), *Evolution and Culture,* The MIT Press, Cambridge, 2006, pp. 133-145.

37 D. Cope, *The Algorithmic Composer,* A-R Editions, Middleton, 2000, y *Virtual Music: Computer Synthesis of Musical Style,* The MIT Press, Cambridge, 2001.

38 En líneas generales, midi es a la música lo que el código ASCII es a la escritura: la *lingua franca* entre los programas de computadora y el mundo externo.

39 Richard Dawkins, *The Extended Phenotype: The Gene as the Unit of Selection,* Oxford University Press, Oxford, 1982.

40 Doug Hofstadter, *Le Ton beau de Marot: In Praise of the Music of Language,* Basic Books, Nueva York, 1997.

41 Tomado, con revisiones, de correspondencia personal, 26 de abril de 2010.

42 Puede leerse completa en D. C. Dennett, «*Homunculi* Rule: Reflections on *Darwinian Populations and Natural Selection* by Peter Godfrey-Smith, Oxford University Press, 2009», *Biology and Philosophy,* 26 (4): 475-488, 2011, y P. Godfrey-Smith, «Agents and Acacias: Replies to Dennett, Sterelny, and Queller», *Biology and Philosophy,* 26 (4): 501-515, 2011.

43 D. C. Dennett, «Memes and the Exploitation of Imagination», *Journal of Aesthetics and Art Criticism,* 48 (2): 127-135, 1990; *Consciousness Explained, op. cit.,* y *Darwin's Dangerous Idea, op. cit.*

44 D. C. Dennett, «The New Replicators», en M. Pagel (comp.), *The Encyclopedia of Evolution,* vol. 1, Oxford University Press, Oxford, 2002, pp. E83-E92; también *Breaking the Spell, op. cit.*

45 R. Dawkins, *The Selfish Gene,* Oxford University Press, Oxford, 1976, p. 206.

[46] Alain, *Propos d'un Normand 1906-1914,* Gallimard, París, 1908.

VII. Herramientas para pensar acerca de la conciencia

Armados con decenas de herramientas de pensamiento, finalmente llegamos al tema que para muchos es el fenómeno más enigmático de todo el universo. En efecto, no poca gente ha sostenido que es irremediablemente misterioso. Nunca *jamás* entenderemos la conciencia, declaran; sistemáticamente desafiará los mayores esfuerzos de nuestra ciencia y nuestra filosofía hasta el fin de los tiempos. Como no hay *buenas* razones para creer en este obstáculo intelectual, tengo que concluir que es pensamiento desiderativo. Algunas personas odian la idea de que revelemos los secretos de cómo trabaja la mente consciente, y sólo para asegurarse de que no les impongamos nuestra comprensión de ellos, aconsejan que lo abandonemos como causa perdida. Si seguimos su consejo, tendrán razón, así que ignorémoslos y emprendamos esta difícil pero no imposible misión.

53. Dos contraimágenes

Muchas de las herramientas que he presentado abordan la mente de una u otra forma —creer, pensar, etc.—, pero pospuse hasta este momento los espinosos problemas de la conciencia. Hay una razón: cuando la gente empieza a cavilar sobre la conciencia, le da por exagerar sus nociones de lo que debe ser la conciencia y por engañarse a sí misma. Afronta los temas más difíciles antes de haber podido ver cuánto del trabajo (y juego) de la mente puede explicarse sin haber resuelto primero los problemas de siempre acerca de la experiencia consciente. Ahora que hemos montado un campamento base, ¿estamos listos para ascender a la cima? Sí, pero si pensamos en el problema de esa manera, ya estamos cometiendo un error de imaginación. Nuestras mentes no tienen una sola cumbre magnífica que sea la conciencia. A diferencia de una tradición que se remonta al menos hasta Descartes en el siglo XVII, los fenómenos conscientes no son ni los fenómenos más «fundamentales» ni los más «elevados» de nuestras mentes.[1] Una mala imagen seductora necesita una contraimagen que la neutralice, así que presento aquí un simple calibrador de imaginación para que empecemos: recuerda la maravillosa canción de Cole Porter «You're the Top» («Eres la cúspide») y piensa que a lo mejor tú *no* eres la cúspide: no la cima de la montaña, sino la montaña entera, y lo que sabes y puedes decir sobre la montaña que eres tú no es la vista desde la cima, sino varias vistas a medio camino. Es buena idea pensar en los fenómenos de la conciencia más bien como el contorno de pelo que hay en la cabeza de un hombre que se está quedando calvo. Ten esta imagen presente.

Y aquí hay otra: la conciencia *no* es un medio, como la televisión, en el que la información pueda transformarse o registrarse, y no hay un lugar en el cerebro donde «todo se reúna» para la apreciación de algún testigo central —llamo a este lugar imaginario el Teatro Cartesiano—.[2] La conciencia es *más como la fama que como la televisión:* fama en el cerebro, celebridad cerebral, una manera en la que algunos contenidos llegan a ser más influyentes y memorables que la competencia. En vez de argumentar a favor de esto, sencillamente te ofrezco esta herramienta de pensamiento (tómala o déjala) con un consejo amistoso: siempre que pienses en la entrada a la conciencia como un arribo al cuartel general o como la traducción de señales de pico neurales a algo más, acuérdate de estas contraimágenes y pregúntate si no estarás imaginando equivocadamente los fenómenos.[3]

54. La corazonada zombi

La mayoría de la gente tiene la corazonada —realmente no pasa de eso— de que ningún robot (hecho de silicio, metal, plástico, etc.) podría jamás ser consciente de la manera como lo somos los humanos. Hay algo en nuestros cuerpos y cerebros vivientes, respiratorios, orgánicos, que es necesario para la conciencia. Es una intuición que prácticamente no necesita bombearse, de tan generalizada que está, y a lo mejor esta gente tiene razón. Sin embargo, ahora que hemos llegado a comprender mejor la manera como nuestros mismos cuerpos y cerebros pueden verse como hechos de robots hechos de robots hechos de robots, y así sucesivamente, hasta llegar al nivel de las neuronas, donde las proteínas motoras y otros nanorrobots avanzan arduamente para hacer que todo el sistema funcione, podemos ver que *a lo mejor* esa corazonada sólo es un producto de una imaginación empobrecida: la gente ha estado pensando en robots que son demasiado simples, por muchos órdenes de magnitud. Una vez un amigo respondió a esto, que ya es una táctica que tengo para iniciar la conversación, tratando de cortarla de raíz: «¡Sencillamente no puedo concebir un robot consciente!» Le contesté que eso era una tontería, que lo que quería decir era que *no quería* concebir un robot consciente. Crees que es estúpido y absurdo tratar de tomarse en serio la idea, pero de hecho es un juego de niños concebir un robot consciente o, es más, un trenecito chu chu consciente (como el del cuento *La pequeña locomotora que sí pudo*), o un árbol de Navidad consciente (recuerda esos cuentos ñoños para niños acerca de pinos solitarios que suspiran por tener un hogar). Cualquiera que haya visto *La guerra de las galaxias* ha pasado cerca de una hora concibiendo a R2D2 y C3PO como conscientes. Lo hemos hecho desde que éramos niños, por lo general «sin darle mayor importancia». No sólo es fácil: es una tendencia casi irresistible cuando nos enfrentamos a algo que actúa —y sobre todo que habla— como ser humano.

Un dato interesante: desde el trabajo pionero del neurocientífico Wilder Penfield en Montreal, allá en la década de 1950, ha habido muchas operaciones que dejaron expuestos los cerebros de pacientes despiertos que podían decir cómo se sentía cuando sus cerebros se estimulaban *aquí* o *allá.* No creo que ningún participante u observador de alguna de estas operaciones haya pensado «¡Dios mío! Esto no es una persona, es un zombi. Tiene que serlo, pues cuando vemos en su interior, sólo encontramos materia gris». No; es demasiado obvio —¡mira!, ¡escucha!— que el paciente está consciente. Y realmente sería igual de obvio si abriéramos el cráneo de alguien mientras platica con nosotros y descubrimos que su cavidad craneal está llena de microchips. Nos daríamos cuenta, quizá para nuestra sorpresa, de que un robot consciente no era sólo fácil de concebir o imaginar, sino real.

Algunos filósofos piensan que tu imaginación estaría jugándote una broma si *te tragaras* estas pruebas «meramente conductuales» de la conciencia y *te precipitaras* a

sacar esa conclusión. «No te lo tragues, no te precipites» podría ser su lema. *Probar* que otra persona es consciente es mucho más difícil, pues la persona podría —al menos es una posibilidad lógica— ser un «zombi». No un zombi al estilo vudú de los que uno ve en las películas o de los que se disfrazan en Halloween. Los muertos vivientes se distinguen sin problemas de las personas normales por su comportamiento (y su aspecto horroroso). En contraste, los zombis de los filósofos pueden ser una compañía muy agradable, el alma de la fiesta, tan afectuosos, dichosos y espontáneos como cualquiera que conozcas. Algunos de tus mejores amigos podrían ser zombis. Los zombis de los filósofos son (por definición) conductualmente indistinguibles de los seres humanos conscientes normales, pero «no hay nadie en casa»: carecen por completo de una vida interna o de cualquier experiencia consciente. Sólo desde fuera dan la apariencia de ser conscientes. Si estás de acuerdo con estos filósofos en que es un tema serio, si te preguntas —dada la posibilidad lógica de los zombis de los filósofos— cómo pudo jamás haber una teoría materialista y científica de la conciencia, entonces se ha apoderado de ti la *corazonada zombi*.[4]

Permítaseme reconocer enseguida que puedo sentir la corazonada zombi tan vívidamente como cualquiera. Cuando lo miro bien, como que *sí* parece que la conciencia debe de ser algo adicional a las cosas que hace por nosotros y las que nos hace a nosotros: alguna clase de resplandor íntimo especial o una condición de *aquí estoy* que estaría ausente de cualquier robot, y es imposible imaginar eso como una «mera» actividad física del cerebro. Sin embargo, he aprendido a no hacerle caso a la corazonada. Creo que es un rotundo error, un fracaso de la imaginación, no la percepción de una necesidad. Sea como sea, convencer a otros de esto no es ninguna tarea fácil, y necesitaremos varias bombas de intuición diferentes para liberarnos de las garras de la corazonada zombi.

Para empezar, podemos orientarnos en esta posibilidad lógica si la comparamos con algunas otras. Es *lógicamente* posible que estés viviendo en la Matrix, y toda la vida que ves a tu alrededor y en la que aparentemente participas sea un espectáculo de realidad virtual pensado para mantenerte tranquilo mientras tu cuerpo real yace inmóvil en alguna clase de tanque de alta tecnología. Es *lógicamente* posible que en realidad no existan los átomos de carbono; lo que a los científicos les parece ser átomos de carbono son en realidad archimillones de diminutas naves espaciales pilotadas por extraterrestres cuya misión en la vida es fingir que son átomos de carbono. Es *lógicamente* posible que el universo entero se haya creado hace aproximadamente 6 000 años, con todos los supuestos fósiles en su lugar, y fotones llegando a raudales como si vinieran de galaxias a años luz de nosotros (es *lógicamente* posible que el mundo se haya creado hace diez minutos, con todo y los supuestos recuerdos de tu pasado instalados en tu cerebro). Esas posibilidades lógicas nos pueden parecer premisas divertidas para la ficción, pero no nos

las tomamos en serio como signos de que nuestra física, nuestra química y nuestra biología tengan que ajustarse o abandonarse. ¿Hay algo que haga la corazonada zombi más sustancial, más digna de consideración? Muchos pensadores serios han pensado que sí.

Quizá la abuela de todas las bombas de intuición concebidas para producir algo como la corazonada zombi haya sido inventada hace cientos de años por Gottfried Wilhelm Leibniz, el filósofo y matemático que comparte con Isaac Newton el crédito de haber inventado el cálculo. Era tan listo e ingenioso como cualquier pensador de su época, y sin embargo se tragó esta bomba de intuición ideada por él mismo.

> Y si se imagina que existe una máquina construida para pensar, sentir, tener percepción, ésta se podrá concebir agrandada, conservando las mismas proporciones, de tal manera que se pueda entrar en ella como si fuera un molino. Supuesto esto, visitándola por dentro no se hallarán más que piezas que se impulsan las unas a las otras, y nunca nada con qué explicar una percepción. *Por tanto,* es en la sustancia simple, y no en la compuesta o en la máquina, donde es necesario buscar la percepción.[5]

Este «por tanto» es uno de los *non sequiturs* más flagrantes de toda la filosofía. Leibniz no interpone ningún argumento a favor de la conclusión: piensa que es demasiado evidente para necesitarlo. Recuerda a William Bateson, el genetista de principios del siglo XX que no podía imaginar que los genes pudieran ser entidades materiales (véase el apartado 22). Tal como Bateson no tenía modo de tomarse en serio la extravagante idea de 3 000 millones de pares de bases en una doble hélice dentro de cada célula (¡ridículo!), Leibniz no tenía modo de tomarse en serio la idea de un «molino» con *billones* de partes móviles. Él sin duda insistiría en que «tan sólo agregar más partes» no podría llevarte de maquinaria a mente, pero eso sería sólo su corazonada, nada que pudiera pretender demostrar. Sin embargo, si Darwin, Crick y Watson han expuesto el fracaso de imaginación de Bateson, Turing ha vuelto obsoleta la bomba de intuición de Leibniz… pero no lo ha hecho, no todavía. En su momento, creo yo, la corazonada zombi caerá poco a poco en el olvido y será vista como un curioso vestigio de nuestro pasado obsesionado con los espíritus, pero dudo que jamás vaya a extinguirse. No sobrevivirá en su forma actual, embrutecedora, pero sí como una mutación menos virulenta, aún psicológicamente poderosa pero despojada de autoridad. Esto ya ha pasado, lo hemos visto. Todavía *parece* como si la Tierra estuviera quieta y el Sol y la Luna giraran a su alrededor, pero hemos aprendido que lo acertado es ignorarlo y tomarlo como una mera apariencia. Todavía *parece* como si hubiera una diferencia entre algo en reposo absoluto y algo que simplemente no se está acelerando dentro de un sistema

inercial, pero hemos aprendido a no confiar en esta sensación. Preveo que llegará el día en que los filósofos, los científicos y los profanos se reirán entre dientes frente a los rastros fósiles de nuestra previa confusión sobre la conciencia: «Todavía *parece* como si estas teorías mecanicistas de la conciencia dejaran algo fuera, pero desde luego que ahora comprendemos que eso es una ilusión. De hecho sí explican todo lo que necesita explicarse acerca de la conciencia».

En realidad, muchos experimentos mentales de filósofos fomentan la lealtad continuada a la corazonada zombi. Uno de ellos es la famosa habitación china de John Searle, que fue la inspiración para que yo acuñara la expresión de «bombas de intuición». Ésta pronto será desmontada frente a tus ojos, pero antes quiero explorar un poco más cuidadosamente el concepto de los zombis filosóficos.

55. Zombis y zimbos

Cuando la gente dice que puede concebir a los zombis (filosóficos), tenemos derecho de preguntar cómo lo saben. ¡Concebir no es fácil! ¿Puedes concebir más de tres dimensiones? ¿La curvatura del espacio? ¿El entrelazamiento cuántico? No basta con tan sólo imaginarlo, y de hecho, nos dice Descartes, eso no es en absoluto concebirlo. De acuerdo con Descartes, para imaginar usas tu cuerpo (en última instancia mecanicista), con todas sus limitaciones (miopía, irresolución, ángulo y profundidad limitados); para concebir usas tu mente, que es un órgano de discernimiento mucho más poderoso, y libre de los requisitos de un mecanismo. Él ofrece un ejemplo convincente de la diferencia: el quiliógono, un polígono regular de mil lados. ¿Puedes concebir uno? ¿Puedes imaginar uno? ¿Cuál es la diferencia? Primero tratemos de imaginar uno. Comienza con un pentágono, digamos; luego imagina un decágono. Es difícil, pero sabes qué hacer: dobla cada lado del pentágono por la mitad y empuja un poco hacia fuera hasta convertir cinco lados iguales en diez. ¿Qué tan lejos debes empujar? Sólo inscribe el pentágono dentro de un círculo y empuja los nuevos lados hasta que intersequen el círculo. Ahora de nuevo, hasta hacer un icoságono, con 20 lados.

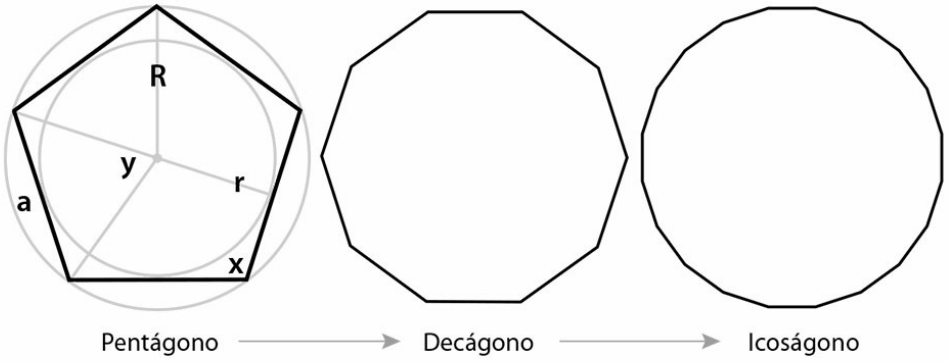


FIGURA 16

Hazlo otra vez, y de nuevo siete veces más, y tendrás un polígono regular de 1 280 lados, que es prácticamente indistinguible de un círculo *cuando lo imaginas,* pero a la hora de *concebirlo* es tan distinto de un círculo —y de un quiliógono— como un cuadrado de un círculo. Si te pido que imagines un círculo dentro de un quiliógono dentro de un círculo dentro de un quiliógono dentro de un círculo, hasta hacer una especie de tiro al blanco, *¿en tu imagen mental* reconoces cuáles son los círculos y cuáles los quiliógonos? No, todos se ven exactamente como círculos, pero no te cuesta ningún trabajo *concebir* la circunstancia en la que te he pedido pensar.

Descartes no te dice que realices esas construcciones; para él la concepción, como la imaginación, es una clase de acto mental directo y episódico, *asir sin tomarse la*

molestia de visualizar, o algo así. De algún modo comprendes (mentalmente) los conceptos pertinentes (LADO, MIL, REGULAR, POLÍGONO), y, ¡*tarán!,* lo tienes. Siempre he visto con sospecha este acto básico de *concebir* cartesiano. Si puedes hacerlo, felicidades, pero me doy cuenta de que yo no puedo (no más de lo que puedo concebir un bebé, en otro sentido del término). No tengo plena confianza de haber *logrado* concebir algo hasta haber manipulado por algún tiempo las ideas pertinentes, probando las implicaciones en mi mente, haciendo ejercicios, de hecho, hasta manejar con soltura las herramientas implicadas (y cuando hago esta gimnasia mental, me recargo mucho en la imaginación, por ejemplo, al explorar en mi cabeza diversos diagramas e imágenes. En pocas palabras, exploto lo que Descartes menospreciaba como mi mera imaginación para conseguir lo que él alabaría como mi concepción). ¿Puedes concebir la teoría de cuerdas? ¿Te parece que es fácil entender todo lo que se dice sobre las tropecientas dimensiones llenas de supercuerdas, «branas» y cosas por el estilo, que es fácil probar si tienen consistencia lógica? Para mí es *ininteligible,* pero por esa mismísima razón no estaría dispuesto a declararla inconcebible o imposible.[6] No estoy convencido, pero tampoco tengo tanta confianza en mis propias aptitudes conceptuales como para descartarla y tildarla de absurda. No he podido concebir la verdad de la teoría de cuerdas *aún.* No deberíamos darles mucho peso a los veredictos que descuidadamente declaran algo concebible o no concebible a falta de demostraciones cuidadosas. Bateson decía que un gen material era «inconcebible», pero si hoy estuviera vivo, fácilmente aprendería a concebirlo. Después de todo, los niños en la escuela primaria aprenden sobre la doble hélice y todos sus travesaños, un fenómeno que demuestran concebir fácilmente una vez que le agarran la onda. Sin embargo, por mucha información nueva y técnicas de imaginación que tengamos, no podremos concebir un cuadrado redondo (un polígono regular de cuatro lados, todos los puntos de cuyo perímetro sean equidistantes de su centro) ni el mayor número primo.

Estoy seguro de que el zombi de los filósofos es conceptualmente incoherente, imposible, una idea fallida. Pero no te fíes de mi palabra. ¿Qué podrías hacer para convencerte de que *puedes* concebir a un zombi de los filósofos? Supón que tratas de imaginar que tu amigo Zeke «resulta ser» un zombi. ¿Qué te convencería o incluso te tentaría de concluirlo?[7] ¿Qué diferencia marcaría toda la diferencia? Recuerda, nada que Zeke pudiera *hacer* debería convencerte de que Zeke es, o no, un zombi. Me doy cuenta de que mucha gente no hace correctamente este ejercicio; es decir, inconvenientemente olvidan o dejan de lado parte de la definición de un zombi de los filósofos cuando intentan su hazaña de concepción. Puede ayudarte a ver si estás cometiendo este error si distinguimos una subespecie especial de zombis a la que llamo zimbos.[8] Todos los zombis tienen sistemas de control no conscientes (desde luego) que obtienen información del mundo (a través de sus globos oculares y oídos zombis) y la explotan para evitar

darse contra la pared, para voltear cuando los llamas, etc. En otras palabras, todos son sistemas intencionales. Pero un zimbo es especial, un zombi que también tiene la suerte de contar con equipo que le permite monitorear sus propias actividades, tanto internas como externas, de modo que tiene estados internos informativos (no conscientes) de orden superior que son *acerca* de sus otros estados internos. Un mayor automonitoreo le permite a un zimbo tener y usar información acerca de esos mismos estados de automonitoreo, y así, indefinidamente. En otras palabras, un zimbo está equipado con *autorrepresentación recursiva:* autorrepresentación recursiva *inconsciente,* si eso tiene algún sentido. Sólo en virtud de este talento especial un zimbo puede participar en la siguiente especie de conversación:

TÚ: Zeke, ¿te caigo bien?
ZEKE: Por supuesto que sí. ¡Eres mi mejor amigo! Tú: ¿Te molestó que te preguntara?
ZEKE: Bueno, sí, fue casi insultante. Me fastidió que me preguntaras.
TÚ: ¿Cómo sabes?
ZEKE: Mmmh. Sólo recuerdo haberme sentido un poco enojado o amenazado, o quizá sólo sorprendido de oír de ti semejante pregunta. ¿Por qué preguntaste?
TÚ: Deja que yo haga las preguntas, por favor.
ZEKE: Si insistes. De hecho esta conversación no me está sentando nada bien.

Recuerda: como un zombi filosófico es declarado conductualmente indistinguible de una persona consciente, conductas como mantener esta conversación forman parte de su repertorio, y a fin de controlarlas va a necesitar autorrepresentación recursiva. Puede «pensar» (a su inconsciente manera zombi) sobre cómo se siente acerca de cómo se sintió acerca de lo que estaba pensando cuando se preguntó si…, y así sucesivamente. Es bastante fácil imaginar que si Zeke se quedara en blanco y se mostrara extrañamente indiferente cuando lo interrogaras así, se despertarían tus horribles sospechas, pero eso sería descubrir que si Zeke es un zombi, no es un zimbo. Siempre deberías asegurarte de que estás pensando en un zimbo cuando preguntas si un zombi de los filósofos es una posibilidad real, pues sólo un ser con autorrepresentación recursiva sabría defenderse en interacciones cotidianas como esta conversación, por no hablar de escribir poesía, formular novedosas hipótesis científicas y actuar en obras de teatro, todo lo cual está entre las competencias de los zimbos, por definición.

A menos que te tomes la molestia de imaginar, en detalle, cuán indistinguiblemente «normal» sería Zeke con respecto al zimbo Zeke, realmente no has tratado de concebir un zombi de los filósofos. Eres como Leibniz: te rindes sin haber hecho el mínimo intento. Ahora plantéate algunas preguntas más. ¿Por qué te *preocuparía* que Zeke fuera un zimbo? O de modo más personal, ¿por qué te preocuparía que tú fueras, o te

volvieras, un zimbo? De hecho, nunca lo sabrías.

¿De verdad? ¿Zeke tiene creencias? ¿O sólo tiene maomeno creencias, «ya sabes, la clase de estados informativos sin conciencia que orientan a los zimbos en sus vidas de la manera como las creencias nos orientan a los demás»? Sólo que aquí, las *maomeno* creencias son exactamente igual de potentes y competentes que las «auténticas», así que éste es un uso indebido del operador maomeno. Esto se pone de manifiesto si imaginamos que los zurdos (como yo, DCD) somos zimbos; ¡sólo los diestros son conscientes!

DCD: ¿Dices que has demostrado que los zurdos somos zimbos? Nunca lo habría imaginado. Pobres de nosotros. ¿En qué aspecto?

DIESTRO: Bueno, por definición no son conscientes. ¿Hay algo peor que eso? que sepa de lo que se pierde. ¿Pero tú qué pretendes al intentar tener una conversación conmigo, un zimbo?

DIESTRO: Bueno, *a mí me parece* que hay alguien ahí.

DCD: ¡A mí también! Después de todo, como zimbo tengo toda clase de competencias de automonitoreo de orden superior. Sé cuando estoy frustrado, cuando algo me duele, cuando me aburro, cuando me divierto, etcétera.

DIESTRO: No. Funcionas *como si* supieras todo esto, pero en realidad no *sabes* nada. Solamente *maomeno* lo sabes.

DCD: Creo que eso es un mal uso del operador *maomeno.* Lo que llamas mi *maomeno* conocimiento es indistinguible de tu supuesto conocimiento *real,* excepto por tu comentario «definicional»: el conocimiento zimbo no es real.

DIESTRO: Pero *hay* una diferencia, ¡*debe haber* una diferencia!

DCD: Eso me suena a llano prejuicio.

Si esto no es suficiente ensayo de cómo sería hacerse amigo de un zimbo, sigue intentando. En serio: piensa en escribir una novela sobre un zimbo atrapado en un mundo de gente consciente, o sobre una persona consciente abandonada en la Isla de los Zimbos. ¿Qué detalles idearías para hacer de ella una historia verosímil? O podrías tomar un camino más sencillo: lee una buena novela y durante tu lectura mantén la hipótesis de trasfondo de que se trata de una novela sobre los zimbos. ¿Qué delata o qué contradice tu hipótesis? Los novelistas pueden elegir puntos de vista o *estilos narrativos,* entre ellos la *narración en primera persona* que usan Herman Melville en *Moby Dick* y J. D. Salinger en *El guardián entre el centeno:*

Llámenme Ismael.

> Cuando algo me preocupa mucho, no nada más me hago guaje. Hasta tengo que ir al baño cuando algo me preocupa… Sólo que no voy. Estoy demasiado preocupado para ir. No quiero interrumpir mi preocupación para ir.

Otros novelistas optan por la *narración en tercera persona omnisciente.* Curiosamente, podría pensarse que el estilo narrativo en primera persona se presta más para sostener la hipótesis zombi. Después de todo, la historia en su totalidad simplemente describe esa *conducta* narrativa de Zimbo Ismael o Zimbo Holden Caulfield. Nosotros sólo vemos sus exteriores y versiones de *lo que afirman ser* recurriendo a sus vidas interiores. Compara esas narraciones en primera persona con, por ejemplo, estas otras en tercera persona de *Persuasión,* de Jane Austen, y *Crimen y castigo,* de Fiódor Dostoievski:

> Ella [Elizabeth] sentía que debían invitar a Mrs. Musgrove y todo su grupo a comer con ellos, pero no podía soportar que la diferencia de estilo, la reducción del servicio que se revelaría en una comida, fueran presenciados por quienes siempre habían sido tan inferiores a los Elliot de Kellynch. Fue una lucha entre el decoro y la vanidad, pero la vanidad se llevó la mejor parte, y Elizabeth de nuevo fue feliz.

> Raskólnikov miró a Sonia y sintió cuán grande era el amor que ella le dirigía. Y, cosa extraña, ese gran amor le produjo de pronto una impresión penosa y dolorosa. ¡Sí, aquello era una sensación extraña y horrible!

Aquí parece que los autores nos dejan «ver directamente adentro de las mentes» de Elizabeth y Raskólnikov, así que ¿cómo podrían ser zimbos? Sin embargo, recuerda que donde la gente consciente tiene un flujo de conciencia, los zimbos tienen un flujo de *inconsciencia.* A fin de cuentas, los zombis no tienen que ser milagrosos; sus conductas se dirigen con una gran cantidad de sucesos internos de enorme complejidad informativa, y se modulan con análogos de emoción funcionales que equivalen a la felicidad, la pena y el dolor. Así, tanto Elizabeth como Raskólnikov podrían ser zimbos, y Austen y Dostoievski usan los términos que todos conocemos y apreciamos de la **psicología vulgar** para retratar sus sucesos internos, tal como los programadores de ajedrez hablan de las «búsquedas» repetidas y los «cálculos» arriesgados de sus programas de computadora. Un zimbo puede sentirse avergonzado por una pérdida de estatus social, o asfixiado por el amor.

Nunca olvides el fracaso de imaginación de William Bateson. Cuando, haciendo mi mayor esfuerzo para evitar esa trampa, busco fisuras en mis suposiciones de trasfondo y mantengo un ojo avizor para encontrar maneras como pudiera demostrarse que estoy

equivocado en lo que respecta a los zombis, siempre doy con descubrimientos imaginarios que muestran, a lo mucho, que hay serias confusiones en torno al concepto de conciencia. Por ejemplo, puedo imaginar que hubiera dos (o siete o 99) clases distintas de supuesta conciencia, y que los zurdos tienen una, los diestros tienen otra, y las langostas tienen otra más. Sin embargo, la única manera como puedo imaginarlo (hasta ahora) es imaginando que *es posible distinguirlos por las siguientes diferencias funcionales:* los zurdos no pueden hacer X, y los diestros no pueden hacer Y, etc. Pero esas diferencias distinguibles sólo confirman que a fin de cuentas no estamos hablando de zombis filosóficos, pues (por definición) no hay diferencias que puedan distinguirse desde fuera entre los zombis filosóficos y la gente «genuinamente consciente». Y hasta ahora nadie ha podido expresar un signo distinguible desde dentro de genuina conciencia que no implique de alguna manera que la persona supuestamente consciente *pueda hacer* algo «mental» que nos convenza (a nosotros y a ella) de que es consciente. Sin embargo, sea cual sea esta diferencia mental, cabe suponer que tendría su homóloga versión falsa en el «flujo de inconsciencia» del zombi. Si no es así, ¿por qué no? Tengo, pues, plena confianza en que toda la idea de un zombi filosófico es una especie de alucinación intelectual, una dolencia que puede dejarse atrás. Inténtalo. Más adelante, en esta misma sección, te proporcionaré más ayuda para esta tarea de autoconvencimiento.

56. La maldición de la coliflor

Te veo zampando ansiosamente una ración de coliflor humeante, cuyo solo olor me da un poco de náuseas, y me descubro a mí mismo preguntándome cómo es posible que disfrutes *ese sabor;* y entonces se me ocurre que para ti la coliflor probablemente sepa (¿o deba saber?) distinto. Parece una hipótesis verosímil, especialmente porque sé que a menudo la misma comida me sabe diferente en distintos momentos. Por ejemplo, mi primer trago de jugo de naranja en el desayuno me sabe mucho más dulce que el segundo si interpongo un bocado de *hot-cakes* con jarabe de maple, pero después de uno o dos sorbos de café, el jugo de naranja vuelve a saber (¿aproximadamente?, ¿exactamente?) como en el primer sorbo. **No cabe duda** (¡ring!) de que está muy bien hablar de *la manera como el jugo le sabe a Dennett en el momento t,* y preguntar si es igual o diferente de *la manera como el jugo le sabe a Jones en el momento t.* Llamemos *qualia* a estas maneras como las cosas pueden parecernos.

Esta «conclusión» parece inocente, pero aquí mismo hemos ya cometido el gran error. El paso final supone que podemos aislar los «qualia» de todo lo demás que pasa —al menos en principio o por mor del argumento—. Lo que cuenta como *la manera como el jugo le sabe a x* puede distinguirse, suponemos, de lo que es un simple acompañamiento, una causa contributiva o una consecuencia de esta manera «central». Uno vagamente imagina tomar tales casos y gradualmente reducirlos a lo esencial, dejando de lado sus residuos comunes, la manera como las cosas se ven, suenan, se sienten, saben, huelen a diversos individuos en diferentes momentos, independientemente de cómo se les estimule o cómo se vean afectados no perceptualmente, e independientemente de cómo estén dispuestos a comportarse o a creer en el futuro. El error no está en suponer que en la práctica podemos algunas veces o siempre realizar este acto de purificación con certeza; se trata del error más fundamental de suponer que hay tal propiedad residual para tomarse en serio, por inciertos que sean nuestros actuales intentos de aislamiento de casos.

Los ejemplos que nos seducen son abundantes en toda modalidad. No puedo imaginar, nunca sabré, por lo visto nunca podría saber, cómo le sonaba Bach a Glenn Gould (a duras penas puedo recuperar en mi memoria la manera como me sonaba Bach a mí cuando era niño). Y no puedo saber, por lo visto, cómo es ser un murciélago,[9] o si, en lo que respecta a los colores, tú ves lo que yo cuando volteamos a un cielo «azul» despejado. Estos ejemplos cotidianos nos convencen de la realidad de estas propiedades especiales —esos sabores, aspectos, olores, sonidos subjetivos— que luego aparentemente aislamos para definirlos en este acto de destilación filosófica. Así nacen los qualia.

Qualia es un término «técnico» para algo que no podría sernos más familiar: la *manera como las cosas nos parecen.* Se diría que no hay nada que pudieras conocer más

íntimamente que tus propios qualia; deja que el universo entero sea una inmensa ilusión, un mero producto del genio maligno de Descartes, y todavía aquello de lo que *está hecho* ese producto (para ti) serán los qualia de tus experiencias alucinatorias. Descartes decía dudar de todo lo que pudiera dudarse, pero nunca dudó de que sus experiencias conscientes tuvieran qualia, las propiedades por las que las conocía o percibía.

Podría decirse que esta definición de qualia es lo suficientemente clara —*la manera como las cosas nos parecen* a nosotros—, pero aunque los qualia así presentados han sido tema de mucho análisis y discusión entre los filósofos, aún no hay un consenso sobre exactamente qué significa o implica el término, técnicamente hablando. Muchos cognitivistas han hecho la benévola suposición de que los filósofos deben de saber de lo que hablan cuando usan este término especial, y lo han añadido a su vocabulario de trabajo, pero eso es un error táctico. Todavía hay una encarnizada controversia sobre lo que son y no son los qualia, más allá de cualquier cuestión empírica. Hace algunos años propuse en un ensayo que el concepto de raíz de los qualia tiene cuatro condiciones.[10] Los qualia son:

1. inefables,
2. intrínsecos,
3. privados y
4. maneras directamente perceptibles en que las cosas nos parecen.

Es decir, son *1:* de algún modo atómicos a la introspección, y por lo tanto indescriptibles («tendrías que haber estado ahí»); *2:* no relacionales, disposicionales o funcionales (a algunas personas el color rojo puede provocarles ansiedad, pero esa disposición subjetiva no es un quale del rojo); *3:* «Tendrías que haber estado ahí, pero no puedes estar ahí; ¡son míos y sólo míos!», y *4:* conoces tus qualia más íntimamente que cualquier otra cosa.

En casi todos los círculos esto sigue considerándose un buen punto de partida, pero como el objetivo de ese ensayo era mostrar que nada podría satisfacer estas cuatro condiciones, se han discutido ampliamente versiones revisadas o mejoradas del concepto, sin que se haya llegado a ningún consenso. No es raro que un término técnico de mucho uso y generalmente respetado tenga múltiples definiciones incompatibles —piensa en «gen» o «especie» en biología y «causa» en casi todas las ciencias—, pero la confusión en torno a «qualia» me da la impresión de ser peor, un caballo de Troya para cualquier persona de otra disciplina que piense en el concepto como un regalo de la filosofía que podría resultar útil en su propia investigación.

En ese ensayo saqué a relucir otras trece bombas de intuición (además de la de la coliflor), y no las repetiré aquí, pues con los años he ideado otras herramientas

probablemente más poderosas para blandir en mi permanente batalla contra la autocomplacencia expresada en una famosa réplica a la invitación a decir exactamente qué son los qualia. Ned Block desestimó esta escandalosa búsqueda «sólo medio en broma» invocando la legendaria respuesta de Louis Armstrong cuando le preguntaron qué era el jazz: «Si tienes que preguntar, nunca lo vas a saber».[11] Esta divertida táctica expresa perfectamente la suposición que es mi blanco. Si triunfo en mi tarea, esta respuesta de Block, que aún hoy se cuela en algunos círculos, se verá tan chocante e insoportable como la jocosa expresión de incredulidad de un vitalista que al enfrentarse con una persona —«una cosa viviente, ¡aunque no te guste!»— dijo dudar de la misma existencia del *élan vital*.

57. Vim: ¿cuánto cuesta en «dinero de verdad»?

Es una opinión común que aunque pudiera construirse un modelo robótico de la visión a color, digamos, que presentara todos los fenómenos conocidos que los humanos experimentamos, como las imágenes residuales complementarias y las ilusiones de contraste de color, y aunque el robot tuviera procesos internos similares que explicaran sus imágenes residuales y demás, como «es sólo un robot» no podría tener qualia rojos y azules. Los estados funcionales que señalaran o representaran cosas de color frente a los ojos de cámara de televisión del robot no tendrían ese algo adicional que nosotros sí. El famoso ensayo de Thomas Nagel «What Is It Like to Be a Bat?» nos ofrece una manera estándar de aludir a los estados conscientes, si los hay, de una entidad.[12] Ser un robot que tiene una imagen residual no *sería como* nada. ¿Por qué tanta gente piensa que esto es obvio? Tal vez porque imaginan un robot relativamente simple y no reparan en que no pueden sacarse conclusiones sobre *todos* los robots a partir de hechos acerca de todos los robots *simples*. Desde luego, si *defines* los qualia como *propiedades intrínsecas* de experiencias consideradas aisladas de todas sus causas y efectos, y lógicamente independientes de todas las propiedades disposicionales, entonces está lógicamente garantizado que eludirán todo análisis funcional. Por mucha ingeniería habilidosa que se tuviera, no podría dotarse de qualia a un robot —pero esto es una victoria pírrica, pues no hay razón para creer que existan tales propiedades intrínsecas—.

Para verlo, compara los qualia de la experiencia con el *valor* del dinero. Algunos estadunidenses ingenuos parecen creer que los dólares, a diferencia de los euros y el yen, tienen un *valor intrínseco*. El turista de la caricatura pregunta: «¿Cuánto cuesta en *dinero de verdad*?», con lo que quiere decir cuánto cuesta en dólares. Empujemos esta idea un poco más lejos: esos estadunidenses ingenuos están dispuestos a intercambiar dólares por euros; estarían encantados de «reducir» el valor de otras monedas a su paridad con los dólares (o a bienes y servicios), pero cuando lo hacen, tienen la sensación de que los dólares son diferentes. Cada dólar, declaran, tiene algo lógicamente independiente de los poderes de intercambio funcional que comparte con todas las otras monedas en circulación. Un dólar tiene cierto no sé qué. Cuando lo contemplas, puedes detectar que tiene un aura de valor, quizá menos que antaño, pero aún perceptible: llamémoslo el *vim* del dólar (del latín *vis,* «fuerza»). Entonces, oficialmente, vim es el valor económico *intrínseco,* no relacional y no disposicional, del dólar. Las libras esterlinas, los euros y otras monedas por el estilo no tienen valor intrínseco: son tan sólo sustitutos simbólicos; son canjeables por dólares, y por lo tanto tienen un valor económico *derivado,* ¡pero no tienen vim! ¡Qué pena para esos europeos! ¡Sus monedas carecen de valor económico intrínseco! ¿Cómo pueden soportarlo? ¿Cómo encuentran motivación para ganar su salario, esos pobres *vimbis*? Los que tenemos la fortuna de que nos paguen en dólares adquirimos artículos con una buena porción de vim incluida. ¡Con

razón tantas naciones han decidido adoptar el dólar estadunidense como moneda oficial! Hasta los extranjeros pueden percibir el vim de un dólar.

Eso dicen nuestros turistas estadunidenses imaginarios. Así definido, está garantizado que el vim de cada dólar eludirá siempre las teorías de los economistas, pues ninguna teoría económica podría jamás explicar el *valor económico intrínseco.* ¿Mucho peor para la economía? La existencia del vim convertiría la economía en una ciencia gravemente incompleta, pero por fortuna no tenemos ninguna buena razón para creer que haya tal cosa como un valor económico intrínseco. El vim evidentemente es un producto de la imaginación, un artefacto de las sinceras corazonadas de esos estadunidenses ingenuos, y podemos explicar el artefacto sin aceptarlo.

Algunos participantes en los debates sobre la conciencia son como los turistas imaginarios. Simplemente insisten, categóricamente, en que sus intuiciones acerca de las *propiedades fenoménicas intrínsecas* son un punto de partida no negociable para cualquier ciencia de la conciencia. Semejante convicción debería considerarse un *síntoma* interesante que amerita un diagnóstico, un dato que cualquier ciencia de la conciencia debe explicar, en el mismo espíritu en que economistas y psicólogos podrían proponerse explicar por qué tanta gente sucumbe a la poderosa ilusión de que el dinero tiene un valor intrínseco. (Cuando Europa cambió al euro, la gente que estaba acostumbrada a concebir los precios en francos, marcos, liras y demás pasaron por un extraño periodo cuando ya no pudieron contar con las «traducciones» a *sus* versiones locales de «dinero de verdad».)[13]

Hay muchas propiedades de los estados conscientes que pueden y deben someterse de inmediato a mayor investigación científica, y cuando tengamos sus informes, bien podremos descubrir que todas nos satisfacen como explicación de qué es la conciencia. Esto es lo que pasó, a fin de cuentas, con el otrora «misterio» de qué es la *vida.* El vitalismo —la insistencia en que hay un grande y misterioso ingrediente extra en todas las cosas vivas, llamado *élan vital*— resulta haber sido un fracaso de la imaginación. Hoy en día el vitalismo está extinto, aunque todavía hay por ahí unos cuantos maniáticos que no se han rendido. Inspirados por esa feliz historia de éxito, ahora podemos seguir adelante con nuestra exploración científica de la conciencia. Si llega el día en que todas las características demostrables de la conciencia sean explicadas, todas las deudas intelectuales reconocidas sean pagadas y aun así sigamos viendo que algo grande *está* faltando (si de verdad es importante, en algún momento debería llamar la atención), quienes tuvieron una corazonada inquebrantable tendrán la oportunidad de exclamar que nos lo dijeron. En el ínterin, pueden preocuparse por cómo eludir el diagnóstico de que, tal como los vitalistas antes de ellos, se dejaron engañar por una ilusión. Así, pues, aquí está el dilema para quienes creen en los qualia como propiedades intrínsecas de las experiencias: ¿cómo distinguen entre su convicción y el error de los estadunidenses

ingenuos? (¿O es que tienen razón estas personas? ¡Los dólares *sí* tienen vim, como cualquiera puede intuir!)

58. El triste caso del señor Clapgras

¿Entonces qué son los qualia, si no propiedades *intrínsecas* de las experiencias conscientes? Una noche, mientras nos tomábamos una excelente botella de Chambertin, el filósofo Wilfrid Sellars me dijo: «Dan, ¡los qualia son lo que hace que la vida valga la pena!» Es una idea atractiva. Consideremos qué implicaría sobre los qualia. Para ver de lo que se trata, presentaré una bomba de intuición contra un telón de fondo de trabajo reciente en neurociencia cognitiva sobre varias patologías singulares y antiintuitivas: la *prosopagnosia* y el *mal de Capgras.*

Las personas aquejadas de prosopagnosia tienen una visión normal en casi todos los aspectos, pero no pueden reconocer caras. Distinguen entre hombres y mujeres, viejos y jóvenes, africanos y asiáticos, pero al estar frente a varios amigos cercanos del mismo género y edad, no pueden reconocer quién es quién hasta que oyen una voz o detectan alguna otra peculiaridad que los identifique. Si a uno de ellos se le presenta una fila de fotografías en las que aparezcan políticos y estrellas de cine famosos, parientes y extraños anónimos, y se le pide que señale a los que conoce, por lo general lo hará al azar. A los que no tenemos prosopagnosia nos puede costar trabajo imaginar cómo es mirar a nuestra propia madre, por ejemplo, y ser incapaces de reconocerla. A algunos les costará creer que exista la prosopagnosia. Cuando le platico a la gente sobre estos fenómenos, a menudo me topo con escépticos que están seguros de que todo me lo estoy inventando. Sin embargo, debemos aprender a tratar esas dificultades como indicadores de nuestra endeble imaginación, no percepciones de imposibilidad. La prosopagnosia (del griego *prosopon,* que significa «rostro», y *agnosia,* «no saber») es una patología bien estudiada, no polémica, de la que adolecen miles de personas.

Uno de los hechos más interesantes acerca de (muchas de) las personas aquejadas de prosopagnosia es que a pesar de su incapacidad de identificar o reconocer caras como una tarea consciente, pueden responder de manera diferente a caras conocidas y desconocidas, e incluso responder de modos que indican que, *sin saberlo ellas,* o *encubiertamente,* pudieron identificar las caras que no pudieron cuando se les pidió. Por ejemplo, ese reconocimiento encubierto se demuestra cuando a alguien con prosopagnosia se le muestran fotos y se le dan cinco nombres candidatos entre los cuales escoger. Escogen al azar, pero su respuesta galvánica de piel —una medida de agitación emocional— muestra una marcada elevación cuando oyen el nombre correcto asociado con la foto. O piensa en una sencilla prueba. ¿Cuál de los siguientes *nombres* son de políticos?: Marilyn Monroe, Al Gore, Margaret Thatcher, Mike Tyson. Deberías poder ejecutar esta tarea con rapidez, pero si te mostraran cada nombre junto con una fotografía que no fuera de esa persona, tu respuesta se retrasaría notablemente. Esto sólo podría explicarse si *en algún nivel* estuvieras realmente identificando las caras, aunque fuera estrictamente irrelevante para la tarea. Parece, entonces, que en el cerebro hay al

menos dos sistemas visuales de reconocimiento facial en gran medida independientes entre sí: el afectado sistema *consciente,* que no puede ayudar a los sujetos en la tarea que se les presenta en el experimento, y el intacto sistema *inconsciente,* que responde con agitación a los nombres y rostros discordantes. Otras pruebas muestran que el sistema afectado está «más arriba» en la corteza visual, mientras que el intacto tiene conexiones con el sistema límbico, «más bajo». Esto simplifica en exceso una historia más sofisticada sobre las variedades de la prosopagnosia y lo que ahora se sabe sobre las áreas del cerebro implicadas, pero será suficiente para nuestros propósitos cuando veamos la aún más extraña patología conocida como mal de Capgras (descrita por el psiquiatra francés Jean-Marie Joseph Capgras en 1923).

La gente que padece el mal de Capgras de pronto empieza a creer que un ser querido —en la mayoría de los casos un cónyuge, un amante o un padre— ha sido sustituido de manera encubierta con un impostor que es su réplica. Quienes sufren de Capgras no están locos; son personas por lo demás bastante normales, que a consecuencia de una lesión en el cerebro de pronto adquieren esta creencia particular, y la sostienen con tal seguridad, a pesar de su extravagancia o su completa improbabilidad, que en algunos casos matan o le provocan serios daños al «impostor», que en realidad es su ser querido. A primera vista debe de parecer simplemente imposible que una lesión en el cerebro tenga precisamente *este* efecto extraño (¿debemos esperar también que haya gente que a consecuencia de un golpe en la cabeza crea que la Luna está hecha de queso?). Sin embargo, el neurocientífico cognitivo Andrew Young notó que había unos rasgos en común, y sugirió que el mal de Capgras era simplemente el «opuesto» de la patología que produce prosopagnosia. En el Capgras, el sistema de reconocimiento facial cortical está a salvo —así es como el paciente engañado reconoce a la persona parada frente a él como la viva imagen de su ser querido—, pero el sistema límbico inconsciente está inhabilitado, y le extrae al reconocimiento facial toda la resonancia emocional que debería tener. La *ausencia* de ese sutil aporte a la identificación es tan terrible («¡Algo falta!») que equivale a un veto al voto del sistema restante que con toda seguridad identifica a la persona cercana: el resultado que surge de ahí es la sincera convicción del paciente de que está viendo a un impostor. En vez de culpar de la discordancia a su propio sistema perceptual defectuoso, los que sufren del mal de Capgras culpan al mundo de una manera tan metafísicamente extravagante y tan improbable que no cabe duda del poder (poder político, en efecto) que el sistema de reconocimiento facial inconsciente, en ellos afectado, tiene en todos nosotros. Cuando el ansia epistémica de este sistema particular se queda insatisfecha, le da un ataque tal que derroca las aportaciones de los demás sistemas.

Haydn Ellis y Andrew Young propusieron esta hipótesis en 1990, y desde entonces Young, el neurocientífico Chris Frith y otros la han confirmado y ampliado. Desde luego,

hay complicaciones en las que no me detendré, pues quiero usar este trozo concreto de neurociencia cognitiva extensora de imaginación para abrir nuestras mentes a una posibilidad más, que aún no se encuentra pero es imaginable. Es el caso *imaginario* del pobre señor *Clapgras,* un nombre que inventé para recordarnos su inspiración: el síndrome real del mal de Capgras (este escenario se suma a una multitud de bombas de intuición de filósofos que exploran trastornos imaginables en la conciencia de una persona que pudieran influir en la naturaleza de los qualia).

El señor Clapgras se gana la vida modestamente como un sujeto de prueba para experimentos psicológicos y psicofísicos, así que no es ningún ingenuo respecto a sus propios estados subjetivos. Un buen día se despierta y en cuanto abre los ojos grita desesperado: «¡Aaaarg! ¡Algo anda mal! Todo el mundo es… extraño, horrible, algo está *mal* en él! ¡No sé si quiero seguir viviendo en *este* mundo!» Clapgras cierra los ojos y se los frota; con cuidado los vuelve a abrir, sólo para enfrentarse una vez más a un mundo extrañamente repugnante, familiar pero también diferente en algún sentido indescriptible. Eso dice él. «¿Qué ves cuando miras hacia arriba?», le preguntan. «Un cielo azul, con unas cuantas nubes blancas algodonosas, unos botones amarillo verdoso en los árboles primaverales, un cardenal rojo brillante posado en una ramita», responde. Al parecer su visión del color es normal, pero sólo para cerciorarse le aplican el test de las cartas de Ishihara, que muestra que no es daltónico, e identifica correctamente unas cuantas decenas de fichas de color del test de Munsell. Casi todos están convencidos de que, sea cual sea la dolencia del pobre señor Clapgras, ésta no afecta su visión del color. Sin embargo, un investigador, el doctor Cromaphil, quiere hacerle algunas pruebas más.

Cromaphil ha estado realizando investigación sobre preferencias de color, respuestas emocionales al color y los efectos de diferentes colores sobre la atención, la concentración, la presión sanguínea, la cantidad de pulsaciones por minuto, la actividad metabólica y gran cantidad de otras sutiles reacciones viscerales. A lo largo de los últimos seis meses ha acumulado una enorme base de datos con las respuestas del señor Clapgras, tanto las idiosincrásicas como las comunes, a todos estos tests, y quiere ver si ha habido algún cambio. Vuelve a hacerle algunas pruebas y observa una regularidad asombrosa: todas las respuestas emocionales y viscerales que Clapgras mostraba ante el azul, ahora las muestra hacia el amarillo, y viceversa. Su preferencia por el rojo sobre el verde se ha invertido, al igual que todas sus demás preferencias de color. La comida le da asco… a menos que coma en la oscuridad. Combinaciones de colores que antes consideraba agradables ahora le parece que desentonan, mientras que las combinaciones de sus «opuestos» ahora le parecen agradables, y así. Sigue identificando como rosa chillante el tono que antes aceleraba su pulso (aunque le parece increíble que alguien pueda llamar chillante a *ese* tono de rosa), pero ahora lo tranquiliza, mientras que su complementario, un tono de verde limón que lo tranquilizaba, ahora lo altera. Cuando

mira pinturas, su trayectoria de movimientos sacádicos —esos veloces brincos que dan nuestros ojos cuando escudriñan algo— ahora son profundamente distintas de sus trayectorias anteriores, que aparentemente estaban dirigidas por sutiles efectos de captar la atención y desviar la mirada que tenían en sus ojos los colores del lienzo. Su capacidad de concentrarse en problemas aritméticos mentales, que hasta ese momento se reducía seriamente cuando lo ponían en una habitación azul brillante, ahora se reduce cuando lo ponen en una habitación amarillo brillante.

En pocas palabras, aunque Clapgras no se queja de ningún problema de visión del color, y de hecho pasa brillantemente todos los tests estandarizados consistentes en nombrar colores y discriminar colores, ha sufrido una profunda inversión de todas sus reacciones emocionales y de atención a los colores. Lo que le ha pasado a Clapgras, les dice el doctor Cromaphil a sus asombrados y escépticos colegas, es sencillo: ha sufrido una completa inversión de los qualia de color, que ha dejado intactos sus meros talentos cognitivos de nivel superior relativos al color: su capacidad de distinguir y nombrar colores, por ejemplo, talentos que un robot sensible a los colores podría tener.

Ahora bien, ¿qué deberíamos decir *nosotros*? ¿Se han invertido los qualia de Clapgras? Como el ejemplo es imaginario, parece que podemos responderlo como queramos, pero los filósofos llevan años tomándose en serio otros ejemplos imaginarios y piensan que de cómo se resuelvan dependen profundas cuestiones teóricas, así que no debemos simplemente desestimarlo. En primer lugar, ¿es éste un ejemplo *posible*? Puede depender de qué clase de posibilidad estemos hablando. ¿Es lógicamente posible? ¿Es fisiológicamente posible? Son preguntas enteramente distintas. Los filósofos han tendido a ignorar esto último y tacharlo de irrelevante para las preocupaciones filosóficas, pero en este ejemplo pueden transigir. No veo manera de argumentar que el caso sea lógicamente imposible. Clapgras se ha descrito como alguien con una extraña combinación de capacidades a salvo y espeluznantes nuevas incapacidades; disposiciones que normalmente están estrechamente unidas, aquí están disociadas de maneras inauditas, pero ¿es su afección más radical en este aspecto que la prosopagnosia o el mal de Capgras? No estoy seguro de que el estado de Clapgras sea siquiera fisiológicamente imposible; hay casos muy estudiados de sujetos que pueden reconocer los colores a la perfección pero no llamarlos por su nombre (anomia de los colores), y de sujetos que se vuelven daltónicos pero no se percatan de su nuevo déficit, y despreocupadamente inventan y nombran colores al azar sin darse cuenta de que están adivinando. Clapgras, igual que quienes sufren del mal de Capgras, no tiene problemas para reconocer o nombrar; es el sutil sabor inefable lo que se ha torcido en él: las disposiciones personales que hacen que valga la pena mirar pinturas, pintar las habitaciones, elegir combinaciones de colores. Los efectos de los colores que contribuyen a que la vida valga la pena es lo que cambió en Clapgras: en otras palabras (si Sellars tenía razón), sus qualia de color.

Supongamos que le comunicamos el asunto a Clapgras y le preguntamos si sus qualia de color se han invertido. Tiene tres respuestas posibles: *Sí, No* y *No sé*. ¿Cuál debería dar? Si comparamos mi historia de Clapgras con los muchos relatos de qualia invertidos que los filósofos han divulgado con seriedad y discutido largo y tendido, la innovación más perturbadora es la posibilidad de que Clapgras pudiera tener sus qualia invertidos y ni siquiera enterarse. Recuerda que el doctor Cromaphil tuvo que proponer esta hipótesis a sus colegas escépticos, y Clapgras bien puede compartir su escepticismo. Después de todo, él no sólo no se ha quejado de ningún problema con sus qualia de color (como en las historias habituales), sino que también se ha convencido de que su visión del color está perfectamente de la misma manera como convenció a los investigadores: pasando fácilmente los tests estandarizados de visión del color. Este rasgo de la historia debería provocar cierta incomodidad, pues por lo general en la bibliografía filosófica se supone que esa clase de autopruebas conductuales son irrelevantes: **no cabe duda** (¡ring!): *esos* tests no tienen ninguna influencia en los qualia. Esos tests comúnmente se caracterizan por no tener absolutamente ningún poder para iluminar o constreñir los dilemas de los qualia. Sin embargo, como mi variación demuestra, las imaginaciones de los filósofos han pasado por alto la posibilidad de que alguien al menos se vea tentado de *contar* con esos tests para salvaguardar su propia confianza en que sus qualia no han cambiado.

¿Pueden tus qualia mantenerse constantes mientras tú experimentas un cambio de «sentimiento»? Las opiniones de los filósofos están divididas respecto a cómo responder esas preguntas relativas a las definiciones de los qualia. Piensa en el efecto del glutamato monosódico, el potenciador de sabores. No cabe duda de que hace que la comida parezca más sabrosa, con un sazón más fuerte, pero ¿*cambia los qualia* de la comida, o simplemente realza la sensibilidad de la gente a los qualia que ya estaba disfrutando? Éste es un llamado a aclarar el concepto de los qualia, no una pregunta empírica sobre la zona de acción del glutamato monosódico, o la variación en respuestas humanas a él según muestren los informes de los sujetos, pues mientras no resolvamos de un modo u otro la pregunta conceptual, cualquier descubrimiento sobre los procesos neurales subyacentes o la heterofenomenología del sujeto será sistemáticamente ambiguo. Lo que quiero saber es simplemente cómo piensan usar la palabra *qualia* los filósofos: ¿identifican todos los cambios en respuestas subjetivas como cambios en los qualia, o hay algún subconjunto privilegiado de respuestas que en efecto anclan los qualia? La idea de cambiar la opinión estética que uno tiene sobre un quale particular —o la manera como responde a él—, ¿es un sinsentido o no? Mientras uno no tome decisiones sobre tales preguntas relativas a su definición, el término no es sólo vago o difuso: es incorregiblemente ambiguo y evasivo frente a dos (o más) ideas fundamentalmente distintas.

¿Se invirtieron los qualia de color de Clapgras? Algunos filósofos dicen que no he dado suficientes detalles al describir su afección. He descrito sus competencias

conductuales —reconoce, distingue y nombra correctamente los colores, mientras que responde «equivocado» en muchos otros aspectos—, pero he evitado describir su estado subjetivo. No he dicho si cuando ve un limón experimenta *verde subjetivo intrínseco* o, digamos, *azul subjetivo intrínseco.* Pero de eso se trata: estoy rebatiendo la suposición de que estos términos nombran propiedades reales de su experiencia. Supongamos que agrego que cuando le preguntan, Clapgras *dice:* «Dado que sigo viendo los limones *como* verdes, desde luego que mi experiencia incluye la propiedad de verde subjetivo intrínseco». ¿Eso resuelve algo? ¿Tenemos alguna confianza en que sabe lo que quiere decir cuando pronuncia estas palabras? ¿Debemos creerle? ¿No estará subyugado por una teoría filosófica que no merece su lealtad?

Ésta es la principal debilidad de los métodos filosóficos que habitualmente se emplean en estos casos: los filósofos tienden a suponer que todas las competencias y disposiciones que presenta la gente normal con respecto a, digamos, los colores, forman un bloque monolítico, invulnerable a la descomposición o a la disociación en subcompetencias y subdisposiciones independientes. Esto convenientemente los excusa de abordar la pregunta de si los qualia deben anclarse a algún subconjunto o a alguna disposición específica. Por ejemplo, los filósofos George Graham y Terry Horgan hablan de «conocimiento directo del carácter fenoménico mismo, conocimiento que proporciona la base experiencial para las capacidades de reconocimiento o discernimiento [de una persona]».[14] ¿Cómo saben que este «conocimiento directo» es la «base» para el reconocimiento o el discernimiento? Cabe suponer que las personas con prosopagnosia tienen conocimiento directo de las caras que ven, o al menos con los «qualias visuales» de esas caras familiares, pero no los pueden reconocer como los qualia que se experimentan cuando ven las caras de sus amigos y parientes. Si, para volver a prestarle atención a Wilfrid Sellars, los qualia son lo que hace que la vida valga la pena, entonces puede ser que los qualia *no* sean la «base experiencial» de nuestra capacidad de reconocer colores día a día, de discernir colores, de llamarlos por su nombre.

59. La baraja afinada

En un famoso artículo, el filósofo David Chalmers distingue entre los problemas «fáciles» de la conciencia y lo que llama el Problema Difícil (con mayúscula) de la conciencia.[15] Lo que Chalmers consideraría «fácil» no deja de ser bastante difícil. Piensa, por ejemplo, en estas preguntas sobre la conciencia que verdaderamente invitan a reflexionar:

1. ¿Cómo la conciencia nos posibilita a hablar acerca de lo que vemos, los sonidos que oímos, los aromas que olemos, etc.? (Simplificando en exceso, ¿cómo las zonas del lenguaje del cerebro usan la información de las zonas perceptoras para formar los informes y respuestas que podemos dar?)
2. Cuando estamos realizando una actividad rutinaria (una que podemos «hacer hasta dormidos»), ¿por qué nuestra conciencia entra en acción cada vez que nos topamos con un problema, y cómo nos ayuda a solucionarlo?
3. ¿A cuántas cosas que se mueven por su cuenta podemos seguirles la pista simultáneamente y cómo lo hacemos? (La respuesta es por lo menos a cuatro; puedes verlo por ti mismo experimentando una alucinante demostración de este fenómeno, conocido como indexación FINST, en <http://ruccs.rutgers.edu/finstlab/mot-movies/mot-Occ-baseline.mov>).
4. Cuando tienes algo «en la punta de la lengua» —cuando sabes que lo sabes y *casi* puedes recuperar la respuesta—, ¿qué es lo que está pasando?
5. ¿Por qué tienes que ser consciente de un chiste para que te parezca gracioso?[16]

Estos problemas son relativamente fáciles, según Chalmers, porque implican las *funciones* cognitivas de la conciencia, las cosas que podemos *hacer,* usando los procesos de dirección de la atención y procesamiento de la información en el cerebro, el rastreo, el recuerdo, el traer a la memoria actividades en las que participamos todo el tiempo que pasamos despiertos. Por difícil que sea idear soluciones prometedoras, las soluciones podrán someterse a pruebas y podrán perfeccionarse con la experimentación, y de hecho estamos haciendo grandes progresos en estos «sencillos» problemas. Podemos construir, por ejemplo, modelos informáticos *relativamente* simples que reproducen estas funciones de manera muy convincente, de modo que podemos estar completamente seguros de que el cerebro las lleva a cabo sin magia ni nada sin paralelo en el resto de la naturaleza. Podría construirse un robot en el que se dieran todos estos fenómenos, si no hoy, en un futuro no muy lejano.

Para Chalmers, el Problema Difícil es el problema de la «experiencia», *de cómo es* ser consciente, el inexpresable, inanalizable *es así* de ser consciente. Un robot podría comportarse *exactamente como si* fuera consciente, respondiendo todas nuestras

preguntas, siguiendo la pista a todas las manchas móviles, sucumbiendo al fenómeno de tener algo en la punta de la lengua y recuperándose de él, riendo en el momento oportuno, e (inconscientemente) desconcertándose o anonadándose en los momentos oportunos, pero de hecho allá adentro en casa no habría nadie. El robot sería un zombi, sin la más pálida sombra de la vida interna que tú y yo, como gente normal y consciente que somos, gozamos.

De acuerdo con Chalmers, tú, amable lector, y yo somos conscientes siempre que estamos despiertos. Un zombi de los filósofos no sabe nada de eso —él nunca está despierto y no tiene vida interna—: sólo *desde fuera parece* ser consciente. Y desde luego que él insiste, convincentemente, en que es tan consciente como tú y yo, y que si al decir esto le ponen un detector de mentiras pasará la prueba de sinceridad…, pero, al ser un zombi, se equivoca. (Los zombis también son indistinguibles de las personas conscientes normales cuando los neurocientíficos investigan sus estados cerebrales internos mediante imágenes por resonancia magnética funcional y otras pruebas por el estilo.) Esto deja claro que distinguir entre una persona consciente y un zombi es efectivamente un problema difícil —si es que es un problema—. Y si *eso* es un problema, explicar cómo puede existir esta diferencia es un problema aún más difícil; es *el* Problema Difícil. Algunos pensamos que el Problema Difícil es un producto de la imaginación de Chalmers, pero otros —muchos, para mi sorpresa— tienen la convicción de que hay o habría una diferencia real entre una persona consciente y un zombi perfecto, y que esto es importante.

Permíteme resumir esta curiosa situación: algunos dudamos de la mera existencia del Problema Difícil, pero otros piensan que debemos estar locos para dudarlo: nada podría ser más evidente, dicen, más inmediatamente intuitivo para cualquier ser consciente, que su conciencia especial, y esta maravillosa propiedad de la que disfrutamos es lo que pone a prueba el entendimiento científico (hasta ahora) y por lo tanto merece ser llamado el Problema Difícil. No hay manera de acercar estas dos posiciones alternativas. Una de las dos está rotundamente equivocada. Durante años he tratado de mostrar que, por tentadora que sea la intuición, es necesario abandonarla. Estoy seguro de que la idea tentadora de que hay un Problema Difícil es sencillamente un error, pero no puedo demostrarlo. O, mejor aún, incluso si pudiera demostrarlo, con frecuencia mi demostración caería en oídos sordos, pues algunos filósofos me aseguran que su intuición a este respecto es firme como una roca, una percepción tan evidente e innegable que ningún argumento podría hacerla temblar, deja tú sacudirla. Así, pues, no cometeré el error táctico de intentar desbancar con argumentos racionales una convicción que está más allá de la razón.

Esta actitud me recuerda la de las sinceras convicciones que a menudo expresan quienes acaban de presenciar una espectacular función de magia. Todos los magos saben

que la gente tiene la tendencia a exagerar en la memoria cualquier buen truco de magia que haya visto. El desconcierto y la impresión del momento amplifican sus recuerdos, y sinceramente y de todo corazón insisten en que han visto algo que va más allá de lo que el mago trató de hacerles creer que veían. Algunas personas tienen grandes deseos de creer en la magia. Recuerda el comentario de Lee Siegel sobre la «magia de verdad» que abordamos en el apartado 22, sobre el tejido mágico: «La *magia de verdad* significa la magia que no es de verdad, mientras que la magia que es de verdad y que de hecho puede hacerse *no es magia de verdad*».

Para muchas personas, la conciencia es «magia de verdad». Si no estás hablando de algo que es supercalifragilisticoexpialidoso, entonces no estás hablando de conciencia, el Misterio Más Allá de Toda Comprensión. El periodista científico Robert Wright expresa de manera concisa esta actitud: «Por supuesto, aquí el problema es la afirmación de que la conciencia es "idéntica" a los estados cerebrales físicos. Mientras más tratan Dennett y sus colegas de explicarme qué quieren decir con esto, más me convenzo de que lo que realmente quieren decir es que la conciencia no existe».[17] Cualquier bolsa de trucos en el cerebro simplemente *no podría* ser conciencia, no conciencia *de verdad.* Pero incluso quienes no cometen este error preventivo tienen a menudo debilidad por exagerar los fenómenos de la conciencia (por eso tuve que dedicar una parte tan importante de mi libro *La conciencia explicada* a desinflar la conciencia —la conciencia de verdad—, reducirla, ponerla en su lugar y mostrar a la gente que los fenómenos no eran tan espectaculares como la mayoría cree. Este ejercicio inspiró después a muchos lectores para bromear con que mi libro debió haberse titulado *La conciencia expulsada* o —como sugiere Wright— *La conciencia negada*). A quienes dudan de que se les podría hacer incurrir desconcertados en una visión inflada de la conciencia, quiero pegarles de refilón, con la esperanza de desvanecer su autocomplacencia al señalarles un exquisito e inquietante paralelo con el mundo de la magia con cartas: la baraja afinada. «Durante muchos años, el señor Ralph Hull, el famoso prestidigitador de Crooksville, Ohio, ha dejado perplejo no sólo al público general, sino también a magos aficionados, gente entendida en trucos de cartas y magos profesionales con la serie de trucos de cartas que tiene el placer de llamar "La baraja afinada".»[18] El truco de Ralph Hull se ve y suena, a grandes rasgos, así:

> Muchachos, tengo un nuevo truco que mostrarles. Se llama «La baraja afinada». Esta baraja de cartas está mágicamente afinada [Hull sostiene los naipes junto a su oreja y los baraja, escuchando atentamente el murmullo]. Por sus vibraciones bien afinadas, puedo *oír* y *sentir* la posición de cualquier carta. Escojan una carta, cualquier carta… [Aquí se le ofrece al público la baraja en forma de abanico u otra, un espectador toma una carta, ve cuál es, y por un camino u otro la regresa a la

baraja.] Ahora escucho la baraja afinada, ¿y qué me dice? Oigo vibraciones reveladoras… [bzzz, bzzz, Hull baraja las cartas junto a su oreja y representa diversas manipulaciones y rituales; al finalizar, presenta la carta del espectador].

Hull representó el truco una y otra vez ante un selecto público de magos, y les daba la oportunidad de adivinar cómo lo hacía. Nadie lo hizo. (Recuerda que una regla fundamental de la magia es nunca repetir un truco frente al público; este gran truco audazmente desacató la regla.) Algunos magos quisieron comprarle el truco, pero él no lo vendía. Años después se lo explicó a su amigo Hilliard, quien publicó la explicación en su libro, que él mismo editó. Esto es lo que Hull tenía que decir sobre su truco: «Durante muchos años he realizado este truco y lo he mostrado a cientos de magos y aficionados, y, hasta donde sé, ninguno de ellos jamás descubrió el secreto… *Los muchachos han estado buscando algo demasiado difícil* [las cursivas son mías]».

Como buena parte de la gran magia, el truco ha terminado antes de que incluso te des cuenta de que empezó. El truco, en su totalidad, está en el nombre, «La baraja afinada», y más concretamente en una palabra: ¡«La»! En cuanto Hull anunció su nuevo truco y dijo el nombre frente a su público entusiasta, el truco había terminado. Después de tenderles la trampa a los espectadores de esta manera tan sencilla, y tras pasar un rato con la evidentemente falsa y equívoca cháchara de las vibraciones y el bzzzbzzz-bzzz, Hull hacía una demostración con el relativamente simple y conocido truco de cartas tipo A (en este punto correré la tradicional cortina del secreto; los detalles mecánicos de la prestidigitación, como verás, no importan). Su público, magos entendidos, se daban cuenta de que probablemente estaría haciendo un truco tipo A, hipótesis que pondrían a prueba mostrándose como espectadores tercos y poco cooperativos a fin de frustrar cualquier intento de hacer el truco tipo A. Cuando adoptaban la actitud recalcitrante adecuada para poner la hipótesis a prueba, Hull «repetía» el truco, pero esta vez ejecutaba un truco de cartas tipo B. Entonces los espectadores se apiñaban y cambiaban impresiones: ¿estará haciendo un truco tipo B? Ponen a prueba *esa* hipótesis adoptando la actitud recalcitrante adecuada para frustrar un truco tipo B, pero él de todas formas hace «el» truco: usando el método C. Cuando ponen a prueba la hipótesis de que les está saliendo con el truco C, él cambia al método D, o a lo mejor vuelve a los métodos A o B, pues su público ya «refutó» la hipótesis de que está usando los métodos A o B. Y así seguía, con decenas de repeticiones, con Hull siempre un paso adelante de sus probadores de hipótesis; se había dado cuenta de que podía explotar el hecho de que siempre podía hacer *un truco u otro* de la reserva de trucos que todos ellos conocían, pero encubriendo que estaba sacando diferentes trucos de la caja de sorpresas con el simple recurso de usar el artículo definido: *la* baraja afinada. Como Hull le explicó a Hilliard: «Cada vez que se representa, la rutina desmiente una o más ideas del

espectador, que invariablemente tarde o temprano renuncia a cualquier nuevo intento de resolver el misterio».

El truco de Hull era introducir una sola palabra común —¡por el amor de Dios!—. Ese modesto monosílabo seducía a su público de expertos y paralizaba sus mentes, con lo que les impedía **pensar afuera de la caja**. Se quedaban atorados en un sistema en el que estaban seguros de que tenían que encontrar un gran truco nuevo, así que no podían ver que su(s) problema(s) no tenía(n) una solución, sino muchas. No consiguieron salirse de la caja.

Entonces, sugiero que David Chalmers —sin querer— cometió la misma hazaña de prestidigitación conceptual al declarar al mundo que ha descubierto «el Problema Difícil». ¿*De verdad* hay un Problema Difícil? ¿O lo que parece ser el Problema Difícil no es más que la gran bolsa de trucos que constituyen lo que Chalmers llama los problemas sencillos de la conciencia? Todos éstos tienen explicaciones mundanas que no requieren ninguna revolución en la física ni que aparezca ninguna novedad. Sucumben, con gran esfuerzo, a los métodos estándar de la ciencia cognitiva.

No puedo demostrar que no haya Problema Difícil, y Chalmers no puede demostrar que lo haya. Tiene a su favor una intuición potente, y si generara alguna nueva predicción asombrosa, o prometiera explicar algo por lo demás desconcertante, tal vez nos uniríamos a él en el intento de construir una nueva teoría de la conciencia alrededor de ella, pero la intuición está sola: es difícil de negar, pero por lo demás es teóricamente inerte.

El inventario de los efectos conocidos de la conciencia es grande y creciente, y va de lo mundano a lo exótico. Es difícil seguirles la pista a todos, así que debemos estar alerta ante la posibilidad de que nos castiguen por un error de aritmética, en efecto, cuando creemos haber cuadrado las cuentas con todos los problemas fáciles y descubrimos un residuo sin explicar. Ese residuo puede haberse ya acomodado, sin que nos diéramos cuenta, en el conjunto de fenómenos humanos para los que ya tenemos explicaciones —o al menos caminos de explicación no misteriosos aún por explorar—. ¿Cómo podríamos cometer este «error de aritmética» y luego pasarlo por alto? Contando doble los fenómenos u olvidando que ya habíamos explicado alguno de ellos y que por lo tanto debimos borrarlo de nuestra lista de «fenómenos por explicar». ¿Es factible que hayamos cometido ese error? Piensa en esto: cuando vimos al pobre señor Clapgras, vimos que *algo* grave le pasaba, pero parecía haber dos maneras claramente diferentes de expresar su dura situación:

A. Todas sus *reacciones* estéticas y emocionales a sus qualia de color se invirtieron (mientras que sus qualia permanecieron iguales).
B. Sus qualia se invirtieron, aunque conservaba la competencia para distinguir, identificar y nombrar colores.

Piensa cómo podría argumentarse (un poco como los magos perplejos tratando de entender cuál es el truco de Hull): «*A* no puede ser, porque la única razón que tenemos para decir que sus qualia permanecieron iguales es que su *conducta de nombrar y discernir* ha seguido igual, pero eso no demuestra nada acerca de sus qualia; todas esas conductas son hechos funcionales (meramente) cognitivos, y los qualia son, desde luego, independientes de ellos. Y *B* no puede ser, porque *sólo* sus reacciones han cambiado; Clapgras no se queja de que ahora los colores *le parezcan diferentes,* sino sólo de que exactamente los mismos colores ya no tienen para él el mismo atractivo que antes. Así, tal vez sus qualia de color cambiaron y tal vez no, y —como verás— ¡no hay manera empírica de saber cuál de las hipótesis es verdadera! ¡Es en verdad un problema difícil!»

Este argumento pasa por alto la posibilidad de que los qualia tratados en *A* y *B* no estén haciendo nada. Tanto en *A* como en *B* vemos que la maquinaria de discernimiento funciona igual que antes, mientras que las reacciones de Clapgras a las entregas de esa maquinaria están invertidas. Los qualia se interponen como una especie de intermediario difícil de definir que se imagina que proporciona la base de materia prima o fundamento de las reacciones emocionales, y parece haber entonces dos lugares donde podría ocurrir la inversión: antes de que los qualia «se presenten» a la maquinaria de apreciación, o después de la «presentación», en la manera como la maquinaria de apreciación responde a esos qualia que se le presentaron. Aquí hay un proceso de presentación de más. Sabemos, por ejemplo, que las reacciones negativas (alarmantes, que provocan miedo) pueden detonarse muy al principio del proceso perceptual, y luego «colorean» todo el procesado posterior de esa entrada perceptual, en cuyo caso podríamos decir que las reacciones emocionales *causan* que los qualia tengan el carácter subjetivo que tenían para Clapgras, y no que (a la inversa) la naturaleza «intrínseca» de los qualia causen o fundamenten las reacciones emocionales. Sin embargo, si ya hemos llegado a las reacciones emocionales (o estéticas o afectivas) a la entrada perceptual, los qualia no tienen más «trabajo» que hacer y, desde luego, un zimbo estaría igual de bajoneado que una persona consciente por la inversión de sus reacciones ante las percepciones.

¿Qué añade la historia sobre la baraja afinada a todas las otras bombas de intuición sobre los qualia? Sólo un ejemplo de la vida real de cómo expertos muy inteligentes y entendidos pueden ser inducidos a crear un problema fantasma simplemente por la manera como se les plantea un asunto. Ha pasado; puede volver a pasar. Y esto aporta una perspectiva novedosa sobre el *impasse* y crea una nueva carga de la prueba: ¿cómo sabes que *no* te has tragado nada parecido a la baraja afinada? No estoy diciendo que esto sea concluyente; sólo sugiero que, con esto, quienes dan crédito a la corazonada zombi deberían preguntarse si de verdad es tan «evidente».

60. La habitación china

En la década de 1970 la inteligencia artificial tuvo una oleada de propaganda que exageró tanto el progreso que había tenido el área como el que se le vaticinaba. ¡Las máquinas pensantes estaban a la vuelta de la esquina! El filósofo de Berkeley John Searle estaba seguro de que a él no lo iban a engañar con ese cuento y tramó un experimento mental para demostrarlo. En 1980 publicó «Minds, Brains and Programs», su famoso experimento mental de la habitación china que pretendía mostrar que una «inteligencia artificial fuerte» era imposible. Definió inteligencia artificial fuerte como la afirmación de que «la computadora debidamente programada tiene literalmente estados cognitivos y que de ese modo los programas explican la cognición humana».[19] Más adelante aclaró su definición: «La computadora digital debidamente programada con las entradas y salidas correctas tendría así una mente exactamente en el mismo sentido en que los seres humanos tienen mentes».[20] El artículo de 1980 apareció en la revista *Behavioral and Brain Sciences* (conocida como *bbs*), el buque insignia de la ciencia cognitiva, que tiene un formato especial: cada número contiene varios «artículos objetivo» acompañados de varias decenas de comentarios a cargo de expertos en el área y una respuesta del autor. Como *bbs* es enérgicamente interdisciplinaria, por lo general estos expertos provienen de una variedad de disciplinas, con lo que se ofrece al público un amplio, provechoso y oportuno espectro de reacciones. Ver si estos otros expertos se toman en serio el artículo objetivo, y cómo, es una gran manera de calcular cómo emplearlo o ignorarlo en tu propio trabajo. También puedes aprender mucho sobre las dificultades de la comunicación interdisciplinaria, con gente muy segura de sí misma que furiosamente habla sin escuchar a los demás, o que participa en una lucha académica por equipos del más alto calibre. El artículo objetivo de Searle provocó una tormenta de fuego de refutaciones, entre ellas la mía, en la que acuñé la expresión «bomba de intuición» con el propósito de mostrar algo que era profundamente engañoso en su experimento mental.[21]

La expresión era mía, pero también debo darle crédito a Doug Hofstadter, pues la acuñación surgió de discusiones con él acerca del ensayo de Searle, que incluimos, con comentarios, en nuestra antología *The Mind's I.*[22] Este experimento mental nos pareció fascinante porque era, por un lado, a todas luces un argumento engañoso y falaz, pero, por otro lado, también era a todas luces convincente y propenso a complacer a las masas. ¿Cómo —y por qué— funcionaba? Lo miramos desde el punto de vista de la *ingeniería inversa,* y Doug planteó la táctica de «girar todas las perillas» para ver qué lo hacía funcionar. ¿La historia resiste las deformaciones, o depende críticamente de detalles que *deberían* ser opcionales?[23]

Eso fue hace más de 30 años, y debo reconocer que la habitación china de Searle ha tenido un enorme atractivo y gran resistencia. Es un clásico, que se presenta

probablemente en miles de cursos de licenciatura, y a la fecha se sigue discutiendo una y otra vez. Yo lo he usado durante años en mis propias clases y he aprendido mucho sobre cómo funciona y sobre cómo mostrarle a la gente lo que está mal en él.[24]

Porque es, como ahora mostraré, una bomba de intuición defectuosa, una **muleta de razonamiento** que puede inutilizar tu imaginación a menos que la manejes con mucho cuidado. Sin embargo, antes de ir a esa delicada tarea, quiero reconocer que quizá muchos de mis lectores estén silenciosamente mirando al cielo, o incluso rezongando. No quieren que inutilice este artefacto; les gusta tanto la conclusión —¡la inteligencia artificial es imposible, fiu!— que sus ojos se ponen vidriosos ante la perspectiva de que los arrastre a lo largo de una crítica meticulosa a un argumento vívido y entretenido que apoya algo que desean fervientemente. Antes respondía a esta reacción con desdén. A los lectores que responden así les encanta el hecho de que una eminencia de Berkeley tenga un argumento famoso que busca mostrar que tienen razón, y están contentos de confiar en su autoridad. Los detalles en realidad no les interesan, sólo la conclusión. ¡Qué manera tan antiintelectual de escabullirse!

Pero entonces me sorprendí haciendo prácticamente lo mismo y recapacité sobre mi duro veredicto. Confieso que la mecánica cuántica siempre me ha parecido un poco repelente, profundamente desorientadora y también, bueno, ¡es algo que preferiría que no fuera cierto! Sé que es asombrosamente exitosa para predecir y explicar muchos fenómenos, entre ellos fenómenos cotidianos como la reflexión y la refracción de la luz, y el funcionamiento de las proteínas en la retina que nos permite ver. Se encuentra en el meollo de la ciencia, pero es famosa por ser difícil de entender, incluso por los expertos. Mis varios intentos de dominar las matemáticas de la mecánica cuántica han fracasado, así que soy un observador, interesado pero completamente incompetente, de las controversias que rodean a su interpretación, aunque esto no me ha impedido desconfiar profundamente de mucho de lo que han dicho de ella expertos supuestamente enterados. Y entonces leí el libro de Murray Gell-Man *El quark y el jaguar. Aventuras en lo simple y lo complejo,* un libro de ciencia para los que, como yo, no son especialistas.[25] Para mi deleite, Gell-Mann adopta un tono sensato y desmitificador, y sencillamente les da una paliza a varias de las declaraciones más discutibles que han ganado aceptación (lee su capítulo «Quantum Mechanics and Flapdoodle» [«Mecánica cuántica y verborrea»], para ver a qué me refiero). Había aquí un físico mundialmente famoso y Premio Nobel apoyando mi prejuicio, usando argumentos que yo entendía. ¡Ésta era la física cuántica para mí! Pero luego caí en la cuenta: ¿de verdad entendía sus argumentos o sólo los **maomeno** entendía? ¿Podía estar seguro de que no estaba nada más prendándome de su retórica? Deseé que no hubiera otros físicos que quisieran arrastrarme de vuelta a lo largo de los tecnicismos para mostrarme que me había dejado engañar por este rechazo acreditado. Me gustaba tanto su conclusión que no soportaba los detalles. También yo

me escabullía.

Pero no del todo. Concienzudamente he tratado de formar un juicio sobre los veredictos de Gell-Mann a la luz de lo que otros han escrito desde entonces (y hasta ahora, todo bien). Sigo abierto a la clara posibilidad de que el «sentido común» de Gell-Mann, que me parece tan atractivo, algún día resulte ser un caso más de fracaso de la imaginación en vez de profunda perspicacia. Les rogaría a los lectores que trataran de adoptar la misma actitud abierta e imparcial hacia mi empeño por desmontar y neutralizar la habitación china de Searle. Procuraré que el trago no sea demasiado amargo.

Allá por 1950, Alan Turing había propuesto lo que según él sería la prueba de fuego de la inteligencia de una máquina, ahora conocida como prueba de Turing, en la que un juez tiene una conversación exploratoria con dos entidades, un ser humano y una computadora, a los que no podía ver y con los que se comunicaba a través de un «teletipo» (piensa: pantalla y teclado). El ser humano trata de demostrarle al juez su genuina humanidad, y la computadora trata de engañarlo y hacerle pensar que ella es el ser humano. Si el juez no puede distinguir la diferencia de una manera fidedigna (el programa de) la computadora habrá pasado la prueba de Turing y no sólo se declarará que es inteligente, sino que tiene «una mente exactamente en el mismo sentido en que los seres humanos tienen mentes», como lo expresó Searle en 1988. Pasar la prueba de Turing iba a ser, a ojos de muchos en el campo, la reivindicación de la inteligencia artificial fuerte. ¿Por qué? Porque (junto con Turing) pensaban que no puedes tener una conversación así sin entenderla, de modo que el éxito de la computadora conversadora sería prueba de su entendimiento.[26] Observa que la prueba de Turing no distingue —no puede— entre un zimbo y una persona «realmente consciente», pues todo lo que una persona consciente puede *hacer,* un zimbo puede hacerlo igual de bien. Como muchos han mencionado, la afirmación de que una entidad que pase la prueba de Turing no sólo es inteligente sino *consciente* va en contra de la corazonada zombi. Podría parecer que esto por sí solo descalifica la prueba de Turing como una buena prueba de la mente, pero antes de emitir un fallo deberíamos *mirar los detalles.* El programa que pase la prueba de Turing podrá tener tan sólo un flujo de *inconsciencia* donde nosotros tenemos un flujo de conciencia, pero acabamos de ver una dificultad que se le plantea a la **corazonada zombi**, que insiste en que esto señala una diferencia real. ¿Es el programa del zimbo sólo **maomeno** consciente? ¿Exactamente en qué dimensión se queda corto? A lo mejor la bomba de intuición de Searle arrojará luz sobre esa pregunta: parece ser un argumento por **reducción al absurdo** para desacreditar la idea de una inteligencia artificial fuerte.

Searle nos invita a imaginarlo encerrado en una habitación, simulando a mano un programa gigantesco de inteligencia artificial, que supuestamente entiende chino. Él estipula que el programa pasa la prueba de Turing al frustrar cualquier intento de un

interlocutor humano de distinguirlo de un genuino entendedor de chino. Searle, que no sabe nada de chino, encerrado en la habitación y manipulando afanosamente las cadenas de símbolos conforme al programa, no por eso obtiene comprensión del chino (obviamente), y no hay nada más en la habitación que sí lo entienda (está vacía, fuera de Searle y los «trozos de papel» en los que están escritas las instrucciones de Searle. Al seguirlas con precisión está «simulando a mano» el programa gigantesco). Si Searle no entiende chino, no cabe duda (¡ring!) de que Searle más trozos de papel no entiende chino. Ergo, no hay comprensión del chino en la habitación china, a pesar de su destreza conversacional, que es lo bastante buena para engañar a hablantes nativos del chino. Searle, como una computadora, identifica los símbolos chinos sólo por su forma; para él y para la computadora no son más que diferentes «garabatos» sin sentido, así que Searle en la habitación china sólo *es* una aplicación del programa de que se trata. Hace su trabajo sin ninguna comprensión del chino, ya sea que se ejecute en silicio o en Searle.

¡Es tan simple y convincente! ¿Es que podría haber algo mal en este experimento mental? Pues sí. Cuando Searle lo presentó en Berkeley, los científicos en computación contestaron con lo que Searle llama la Réplica de Sistemas (Berkeley):

> Si bien es cierto que la persona que está encerrada en el cuarto no entiende la trama, el hecho es que ella simplemente forma parte de todo el sistema, y el sistema sí entiende la trama. La persona tiene enfrente un gran cuaderno en el que están escritas las reglas, tiene mucho papel para borradores y lápices para hacer cálculos, tiene «bases de datos» de conjuntos de símbolos chinos. Ahora bien, la comprensión no se atribuye al simple individuo; más bien se atribuye a todo este sistema del que forma parte.[27]

Searle dice algo muy revelador en el transcurso de su respuesta a esta réplica:

> De hecho me siento un poco avergonzado […] porque para empezar la teoría me parece de lo más inverosímil. La idea es que si bien una persona no entiende chino, de alguna manera la *conjunción* de esa persona y unos trozos de papel podrían entender chino. No es fácil para mí imaginar cómo alguien que no esté subyugado por una ideología pudiera encontrar la idea mínimamente verosímil.[28]

Lo que a Searle le parece tan inverosímil no es otra cosa que la intuición fundamental que tuvo Turing cuando creó la idea de un programa de computadora almacenado. Toda la competencia está en el *software.* Recuerda que la máquina de registro del apartado 24 no entiende nada de aritmética, pero la máquina de registro en conjunción con el *software* hace una aritmética perfecta. La unidad de procesamiento central de tu

computadora portátil no sabe nada de ajedrez, pero cuando ejecuta un programa de ajedrez, te puede ganar en un juego, y lo mismo con todas las magníficas competencias de tu computadora portátil. Lo que Searle describe como una ideología está en el centro mismo de la ciencia de la computación, y su solidez se demuestra en todos los ámbitos de la vida. La manera de reproducir la competencia humana y en consecuencia (tarde o temprano) la comprensión es apilar máquinas virtuales encima de máquinas virtuales encima de máquinas virtuales: el poder está en el sistema, no en el *hardware* subyacente. La «extraña inversión de razonamiento de Darwin» resuena agradablemente en la extraña inversión de razonamiento de Turing: mientras que solíamos pensar (antes de Turing) que la competencia humana tenía que derivarse de la comprensión (esa misteriosa fuente de inteligencia), ahora nos damos cuenta de que la comprensión misma es un efecto creado (borboteando) a partir de una gran cantidad de competencias apiladas sobre otras competencias.[29]

Los detalles importan. Searle nunca le dice a su lector a qué nivel está simulando a mano el gigantesco programa de inteligencia artificial. Ésta es su explicación de lo que pasa cuando se ocupa del programa:

> Ahora supóngase que después de esta primera pila de escritura china [la entrada] se me da una segunda pila de comandos en chino junto con un conjunto de reglas para correlacionar la segunda pila con la primera. Las reglas están en inglés, y entiendo estas reglas tan bien como cualquier otro hablante nativo del inglés. Éstas me permiten correlacionar un conjunto de símbolos formales con otro […] Ahora supóngase también que se me da una tercera pila de símbolos chinos junto con unas instrucciones, de nuevo en inglés, que me permiten correlacionar elementos de esta pila con las dos primeras, y estas reglas me enseñan cómo devolver algunos símbolos chinos con ciertas clases de formas en respuesta a ciertas clases de formas que se me dieron en la tercera pila.[30]

Él contrasta esta «correlación» entre «pilas de símbolos» con lo que pasa cuando en vez de eso se introduce una oración o un relato en inglés, y él responde en su inglés nativo. «Desde el punto de vista externo —desde el punto de vista de alguien que lee mis "respuestas"— las respuestas a las preguntas en chino y a las preguntas en inglés son igualmente buenas. Sin embargo, en el caso chino, a diferencia del inglés, presento las respuestas mediante la manipulación de símbolos formales sin interpretar.»[31] ¡Qué contraste! Pero mira lo que Searle sigue dejando fuera. Sabemos que obtiene su «conjunto de reglas» (instrucciones) en inglés, pero ¿son éstas del tipo «suma el contenido del registro 39021192 al contenido del registro 215845085» (código máquina) o del tipo «Define una constante: fila-tamaño y fíjala en 100» (**código fuente**)? ¿Está en

el sótano haciendo operaciones aritméticas a una velocidad vertiginosa (billones de operaciones por segundo) o está siguiendo el código fuente de un programa aplicado muchas capas más arriba? Y si está siguiendo el código fuente, ¿llega a leer algún comentario? Mejor que no, pues no son parte oficial del programa y le darían muchas pistas sobre lo que está pasando («Esto analiza la oración: arroja sustantivos, pronombres, verbos y modificadores, y los identifica como interrogativos, declarativos, imperativos o interjecciones», y luego, unos cuantos miles de millones de operaciones después, «juego de palabras descubierto, cambia a modo conversación…», y luego un conjunto más detallado de miles de millones de operaciones, en el que se pulen las referencias, se evalúan las respuestas alternativas un poco como diferentes jugadas de ajedrez, y finalmente se genera una oración de salida). Si el programa que Searle está simulando a mano puede mantener una admirable conversación en chino, tendrá que consultar enormes bases de datos, no sólo de «conjuntos de símbolos chinos», como dice él, sino de conocimiento cotidiano que comparten los hablantes de chino, y eso es lo de menos. Cuando Searle simula a mano, ¿se da alguna idea de toda esta actividad cognitiva en capas que está ocurriendo, o para él sólo es un montón de aritmética? Piensa cómo podría manejar Searle la siguiente pregunta en inglés:

> Imagina que tomas una letra D mayúscula y la volteas 90 grados en sentido contrario a las manecillas del reloj. Ahora ponla arriba de una J mayúscula. ¿Qué clase de clima te recuerda?

Ahora imagina que se le da a Searle una tarea análoga en chino cuando está concentrado en la habitación china.

> El 4 de junio de 2012 se bloqueó el siguiente posteo de caracteres en Sohu Weibo [un servicio chino de bitácoras de internet]. ¿Puedes descifrar por qué?
>
> 占占占占人 占占占点 占占点占 占点占占 点占占占 灬占占占占

Éstos son caracteres chinos reales (los «garabatos» de Searle), pero la secuencia es una absoluta incoherencia. ¿Por qué la bloquearían las autoridades? Porque el 4 de junio es el aniversario de la masacre en la plaza de Tiananmén, en la que el ejército mató a cientos de protestantes («4 de junio» es tan evocador para los chinos como «9/11» para los estadunidenses). Las imágenes más famosas que surgieron son de un solo hombre valiente que les hizo frente a los tanques. Puedes verlo («人» significa persona) enfrentándose a cuatro tanques a la izquierda, que luego pasan encima de él, y encima de él, y encima de él, y encima de él, y luego se van por la derecha.[32]

La habitación china debería «captarlo», pero a menos que Searle tuviera acceso a los comentarios del código fuente, no se enteraría, pues nunca habría sido evidente para él que al seguir las reglas se estaba elaborando y manipulando una «imagen mental», y luego usando el resultado como un sondeo de memoria. Esto es, antes de responder la pregunta en chino el *sistema* pasaría por un conjunto de actividades que son asombrosamente análogas a las actividades por las que Searle pasó a sabiendas para responder la pregunta en inglés. Podría decirse que el sistema tiene una mente propia, que Searle ni imagina, dándole duro en el cuarto de máquinas.

Cualquier programa que pudiera pasar la prueba de Turing *tendría* que pasar por operaciones «mentales» que imitaran muy bien las operaciones mentales por las que pasamos en el transcurso de una conversación. Supongamos, por ejemplo, que el interrogador en la prueba de Turing empezara a darle al candidato clases de física cuántica con el método socrático de pregunta y respuesta, y que le pusiera al estudiante unos sencillos ejercicios. Searle, en el cuarto de máquinas, estaría obligado a llevar al sistema a través de intrincados ejercicios intelectuales buscando retrasar el fin de su conversación, pero Searle terminaría la dura prueba sabiendo tan poca física cuántica como cuando entró. En contraste, el sistema ahora tendría mucho más conocimientos básicos del área que antes de haber empezado la prueba de Turing, porque habría *hecho los ejercicios*. Este caso particular de la prueba de Turing instalaría una nueva máquina virtual en el programa: la máquina de física cuántica simple.

Tales hechos se oscurecen completamente en la imagen de Searle de los «trozos de papel» y las «reglas» que le permiten «correlacionar» símbolos chinos. No estoy diciendo que él deliberadamente haya ocultado la complejidad del programa que imaginó que simularía a mano, sino sólo que ignoraba las implicaciones de la complejidad. Si piensas en el programa como un montón de reglas relativamente simple, entonces, como Bateson cuando piensa en «partículas de cromatina, indistinguibles unas de otras y de hecho casi homogéneas según cualquier prueba conocida», es probable que los supuestos poderes de comprensión de los programas, como los poderes del ADN, te parezcan «inconcebibles».

Mira lo que acabamos de hacer. Hemos girado la perilla de la bomba de intuición de Searle que controla el nivel de descripción del programa que se sigue. Siempre hay muchos niveles. En el más alto, los poderes de comprensión del sistema no son inimaginables; incluso tenemos cierta percepción de precisamente cómo el sistema llega a entender lo que hace. La respuesta del sistema ya no parece vergonzosa; parece evidentemente correcta. Eso no significa que la inteligencia artificial de la clase que Searle estaba criticando realmente consiga un nivel de competencia que merezca llamarse comprensión, ni que esos métodos, ampliados de las maneras que entonces imaginaron esos investigadores de inteligencia artificial, probablemente hubieran conducido a estas

competencias superiores, sino sólo que el experimento mental de Searle no consigue lo que dice: demostrar la rotunda imposibilidad de una inteligencia artificial fuerte.

Hay otras perillas que girar, pero esa tarea se ha cumplido con creces en la inmensa cantidad de bibliografía a la que ha dado lugar la habitación china. Aquí me concentro en la herramienta de pensamiento misma, no en las teorías y proposiciones a las que estaba dirigida, y muestro que es una herramienta defectuosa: convence empañando nuestra imaginación, en vez de saberla aprovechar.

61. La caída del teleclón de Marte a la Tierra

Ves a la luna salir por el este. Ves a la luna salir por el oeste. Ves dos lunas acercándose una a otra a través del frío cielo negro, y pronto se cruzan y siguen su camino. Estás en Marte, a millones de kilómetros de casa, protegida del frío matador (sin escarcha) del rojo desierto marciano gracias a frágiles membranas de tecnología terrestre: estás protegida pero te has quedado tirada, pues tu nave espacial está descompuesta y no hay modo de repararla. Nunca jamás podrás volver a la Tierra, a los amigos, la familia, los lugares que dejaste atrás.

Pero a lo mejor hay esperanza. En el compartimento de comunicaciones de la inutilizada nave encuentras un teletransportador teleclón Mark IV e instrucciones para usarlo. Si enciendes el teletransportador, ajusta su señal con el receptor teleclón en la Tierra, y entra en la cámara de envío; el teletransportador suave e indoloramente desarmará tu cuerpo y producirá un plano molécula por molécula que se transmitirá a la Tierra, donde el receptor, con su depósito bien surtido de los átomos necesarios, casi instantáneamente, a partir de las instrucciones transmitidas, ¡te producirá a ti!, devuelta a la Tierra a la velocidad de la luz, hacia los brazos de tus seres queridos, que pronto escucharán absortos el relato de tus aventuras en Marte.

Una última inspección de la nave espacial averiada te convence de que el teleclón es tu única esperanza. Con nada que perder, armas el transmisor, enciendes los interruptores indicados y entras en la cámara. Cinco, cuatro, tres, dos, uno, ¡flash! Abres la puerta frente a ti, sales de la cámara de recepción del teleclón y entras a la soleada atmósfera familiar de la Tierra. Has llegado a casa, no tan desmejorada a pesar de la caída de larga distancia desde Marte por el teleclón. Haberte salvado por un pelo de un terrible destino en el planeta rojo amerita una celebración. Mientras tu familia y tus amigos se reúnen en torno a ti, notas cómo todos han cambiado desde la última vez que los viste. Después de todo han pasado casi tres años, y todos han envejecido. Ves a tu hija Sara, que ahora debe de tener ocho años y medio. Te descubres a ti misma pensando: «¿Puede ella ser la niñita que se sentaba en mi regazo?» Por supuesto que sí, reflexionas, aunque debes admitir que más que reconocerla estás extrapolándola del recuerdo y deduciendo su identidad. Está mucho más alta, se ve mucho mayor, y sabe mucho más. De hecho, la mayoría de las células que ahora están en su cuerpo no estaban ahí la última vez que posaste los ojos en ella. Sin embargo, a pesar del crecimiento y el cambio, a pesar del remplazo de células, sigue siendo la misma personita a la que le diste un beso de despedida tres años atrás.

Y en eso se te ocurre: «¿Realmente soy la misma persona que le dio un beso de despedida a esta niña hace tres años? ¿Soy la madre de esta criatura de ocho años o en realidad soy un ser humano totalmente nuevo, de apenas unas horas de edad a pesar de mis recuerdos —o aparentes recuerdos— de días y años atrás?» ¿La madre de esta niña

murió recientemente en Marte, y fue desarmada y destruida en la cámara de un teleclón Mark IV?

¿Morí en Marte? No, claro que *yo* no morí en Marte, pues estoy viva en la Tierra. Con todo, quizá *alguien* murió en Marte: la madre de Sara. Entonces yo no soy la madre de Sara. ¡Pero debo serlo! Todo el propósito de meterme al teleclón era volver a casa con mi familia. Pero sigo olvidando; quizá *yo* nunca entré en ese teleclón en Marte. Quizá era alguien más, si es que eso de verdad ocurrió.

¿Es esa máquina infernal un tele*transportador*, un medio de transporte, o, como el nombre de la marca sugiere, una especie de fabricante de gemelos asesino? ¿Sobrevivió la madre de Sara a la experiencia con el teleclón o no? Ella pensó que sobreviviría. Entró a la cámara con esperanza y expectativas, no con resignación suicida. Su acto fue altruista, por cierto: estaba tomando medidas para proporcionarle a Sara un ser querido que la protegiera; pero su acto también fue egoísta: estaba saliendo de un aprieto y cambiándolo por una situación agradable. Al menos eso parecía. «¿Cómo sé *yo* que eso parecía? Porque estaba *ahí*; yo *era* la madre de Sara teniendo esos pensamientos. *Soy* la madre de Sara. Al menos eso parece.»

Sin lugar a dudas, una canción, un poema o una película pueden teletransportarse. ¿Es un yo la clase de cosa —una cosa «hecha de información»— que puede teletransportarse sin pérdida? ¿Será nuestra renuencia a admitir la teletransportación de personas un poco como la resistencia anacrónica, recientemente superada en casi todos los ámbitos, a los documentos con firmas legales escaneadas electrónicamente? (En 2011 me enteré de que la Sociedad de Miembros de la Universidad de Harvard no aceptaba una firma escaneada en mi carta de recomendación; requerían un poco de tinta seca que realmente hubiera sido ahí depositada por el movimiento de mi mano, y me tomó medio día de paseos en taxi en Beirut obtener el formato, firmarlo y enviarlo de regreso por correo exprés —en papel bond acremado—. Tengo entendido que la Sociedad ya cambió sus criterios, pero espero que Harvard siga empeñándose en poner sellos de lacre en sus diplomas. Hay un lugar para la tradición, en toda su gloriosa gratuidad.)

62. El yo como centro de gravedad narrativa

¿Qué es un yo? Los filósofos llevan siglos lidiando con esta pregunta. El concepto cristiano de un alma inmortal, inmaterial e inexplicable ha cautivado a los pensadores y desviado la investigación seria durante cientos de años, pero todos los días pierde partidarios. La idea de una como cosa mental que se va al Cielo cuando alguien muere es cada día más incoherente. El pensamiento desiderativo que nos impide desecharlo sin más, junto con las brujas y los duendes, es demasiado evidente. Los materialistas que estamos seguros de que la mente es el cerebro (bien entendido) tenemos que afrontar la pregunta de por qué *parece* que cada uno de nosotros tiene tal como cosa mental, o mejor aún: que cada uno de nosotros *es* esa como cosa mental que habita un cuerpo, y más específicamente un cerebro. ¿Simplemente encontramos nuestros yos cuando miramos adentro?

Es bien conocida la manera como David Hume menospreció esta idea en 1739:

> En lo que a mí respecta, siempre que penetro más íntimamente en lo que llamo *mí mismo* tropiezo en todo momento con una u otra percepción particular, sea de calor o frío, de luz o sombra, de amor u odio, de dolor o placer. Nunca puedo atraparme a *mí mismo* en ningún caso sin una percepción […] Si tras una reflexión seria libre de prejuicios hay alguien que piense que él tiene una noción diferente de *sí mismo,* tengo que confesar que ya no puedo seguirle en sus razonamientos. Todo lo que puedo concederle es que él puede estar tan en lo cierto como yo, y que ambos somos esencialmente diferentes en este particular. Es posible que él pueda percibir algo simple y continuo a lo que llama su *yo,* pero yo sé con certeza que en mí no existe tal principio.[33]

El reconocimiento medio en broma de Hume de que otros pueden ser diferentes encuentra eco hoy en día entre quienes se preguntan si yo, por ejemplo, soy un zombi (un zimbo, desde luego), que inocentemente extrapola su propia experiencia empobrecida a los demás. Es una conjetura divertida, pero no creo que nadie se la tome en serio.

Está claro qué no es un yo. No es parte del cerebro, como lo son la amígdala o el hipocampo. Los lóbulos frontales tienen un papel fundamental para evaluar situaciones, intenciones, percepciones y demás, pero no creo que nadie haya cometido el error de *ubicar el yo* en los lóbulos frontales. (Una lobotomía prefrontal es una cirugía atroz, que deja a alguien literalmente como la «sombra de lo que fue», pero no es una yotomía. Como dice el viejo chiste, preferiría tener frente a mí una botella gratis que una lobotomía prefrontal, pero en cualquiera de los casos estaría ahí para vivirlo.)[34] ¿Entonces qué podría ser el yo? Propongo que es la misma *clase* de cosa que un centro

de gravedad, una abstracción que, no obstante su cualidad abstracta, está estrechamente acoplada al mundo físico. Igual que todos los demás objetos materiales, tú tienes un centro de gravedad (sería más preciso decir un centro de masa, pero ignoremos por el momento esas sutilezas). Por ejemplo, si por la distribución del peso de tu cuerpo tu centro de gravedad está más arriba que en el promedio de la gente de tu estatura, seguramente te costará más trabajo estar erguido, y así. Hay muchas maneras de ubicar tu centro de gravedad, que, dependiendo de factores como los zapatos que llevas puestos y la hora de tu última comida, se mueve alrededor de un área más bien pequeña en medio de tu cuerpo. Es un punto matemático, no un átomo o una molécula. El centro de gravedad de un trozo de tubo de acero no está hecho de acero; es más, no está hecho de nada. Es un punto en el espacio, el punto de la línea media que corre por el centro del tubo que sea equidistante de los extremos (a grandes rasgos, dependiendo también de imperfecciones y otras cosas).

El concepto de un centro de gravedad es una herramienta mental muy provechosa por derecho propio. De hecho, saca un promedio de todas las atracciones gravitacionales entre todas las partículas de materia de una cosa y todas las partículas de materia del planeta, y nos dice que podemos reducir todo eso a dos puntos —el centro de la Tierra (su centro de gravedad) y el centro de gravedad de la cosa— y calcular el comportamiento de la cosa en condiciones variables. Por ejemplo, si el centro de gravedad de una cosa en cualquier momento cae fuera de todos los puntos de su base de apoyo, perderá el equilibrio. Por supuesto, teníamos una comprensión intuitiva de los centros de gravedad mucho antes de que Newton entendiera la gravedad («¡Siéntate! Vas a hacer que se vuelque la lancha».) Ahora podemos explicar en detalle cómo y por qué funciona el concepto, y si estamos diseñando vehículos o lámparas de pie, por ejemplo, el propósito de bajar un centro de gravedad, o moverlo a una ubicación más efectiva, muestra que el concepto es casi indispensable en muchas de nuestras actividades. Podrá ser una «ficción de teórico», pero es una ficción muy valiosa de la que pueden generarse gran cantidad de predicciones verdaderas. ¿Puede una entidad abstracta como ésa, que no tiene existencia material, realmente *causar* algo? No directamente, pero las explicaciones que citan un centro de gravedad compiten con las explicaciones que son claramente causales. ¿Por qué esa taza de café no se volcó cuando el velero se inclinó tan drásticamente? «Porque tiene un centro de gravedad inusitadamente bajo» compite con «Porque está pegada a la cubierta».

Podemos decirle a un centro de gravedad una *ficción* de teórico porque comparte con los personajes ficticios la curiosa propiedad de la indeterminación de propiedades. Sherlock Holmes, según lo describe Arthur Conan Doyle en los relatos policiacos de Sherlock Holmes, tiene muchas propiedades, pero donde Conan Doyle guardó silencio no hay verdad alguna. Podemos extrapolar un poco: el autor nunca menciona que

Sherlock tenga tres orificios nasales, así que tenemos derecho a suponer que no los tenía.[35] También podemos estar de acuerdo en que no era un bígamo con una esposa en París y otra en Nueva York. Sin embargo, para muchas preguntas así no hay respuesta. ¿Tenía un lunar en el omóplato izquierdo? ¿Era primo hermano de Oscar Wilde? ¿Tenía una casa rural en Escocia? Estas preguntas y archimillones más acerca de cualquier ser humano *real* deben tener respuestas verdaderas, aunque nunca podamos descubrirlas. No para Sherlock; por ser ficticio, sólo tiene propiedades que su autor dice o insinúa que tiene. Un lector cándido que pensara que los relatos de Sherlock Holmes son verdaderos podría preguntarse si el revisor del tren a Aldershot era más alto o más bajo que Holmes, pero si entiendes lo que es la ficción sabrás que mejor ni lo preguntas. Lo mismo puede decirse de los centros de gravedad. Si te preguntabas si podía ser que a la larga «resultaran ser neutrinos», no estarías comprendiendo su sentido como ficciones de teórico.

¿Entonces qué es un centro de gravedad *narrativa*? Es también una ficción de teórico, postulada a fin de unificar y entender una colección, que de otro modo sería desconcertantemente compleja, de acciones, declaraciones, inquietudes, quejas, promesas, y así sucesivamente, que componen una persona. Es el organizador del nivel personal de la explicación. Tu mano no firmó el contrato, sino tú. Tu boca no dijo la mentira, sino tú. Tu cerebro no se acuerda de París, sino tú. Tú eres el «dueño del registro» del cuerpo viviente que reconocemos como tú (como se dice: es tu cuerpo y tú haces con él lo que quieres). Así como podemos simplificar todas las atracciones gravitacionales entre todas las partes del mundo y un obelisco plantado en el suelo reduciéndolas a dos puntos, el centro de la Tierra y el centro de gravedad del obelisco, podemos simplificar todas las interacciones —los apretones de manos, las palabras que se pronuncian, el garabato de tinta y mucho más— entre dos yos, el vendedor y el comprador, que acaban de concluir una transacción. Cada yo es una persona, con una biografía, un «telón de fondo» y muchos proyectos en curso. A diferencia de los centros de gravedad, los yos no sólo tienen trayectorias a través del espacio y el tiempo: van recogiendo sobre la marcha, acumulando recuerdos, concibiendo proyectos y expectativas.

Hay partes del telón de fondo de las que cada persona probablemente desearía renegar, pero lo que está hecho está hecho; esa parte de la narrativa no puede corregirse. Puede, sin embargo, reinterpretarse a la luz de posteriores elementos biográficos. «Estaba fuera de mí cuando hice eso» es un dicho común, y muchas veces es acertado mostrarse tolerante con esta afirmación que aparentemente es un contrasentido. Lo que la persona quiere decir es que estaba en un estado extremo, un estado no típico de ella, y nos suplica que no la juzguemos en el presente o extrapolemos ese comportamiento al futuro. A veces esto puede hacerse, a veces no. Otra jugada conocida: «Y bien, si tú no

hiciste eso, ¿quién fue?» «Me tentó el diablo.» Por otra parte, a menudo aceptamos esta afirmación, no en sentido literal, sino como una sincera abjuración del carácter y las motivaciones que guiaron la acción. Esto plantea problemas sobre la responsabilidad y el libre albedrío que abordaremos en el siguiente capítulo. Por el momento, sólo observa cómo necesitamos el concepto del yo para trazar una línea (por arbitraria que muchas veces resulte ser) entre lo que tú *haces* y lo que te pasa a ti.

Todo objeto físico tiene un centro de gravedad y todo cuerpo humano viviente tiene un yo, o más bien, todo cuerpo humano viviente tiene un dueño, que es un yo, una especie de administrador que vive en su lugar de trabajo. ¿Sólo un dueño? ¿Podría un cuerpo ser compartido por dos o más yos? La enfermedad conocida como trastorno de identidad disociativo (que antes solía llamarse trastorno de personalidad múltiple), según parece, es un ejemplo de muchos yos que comparten un solo cuerpo, con un yo dominante (el «huésped») y un grupo de «álter». Digo «según parece» porque hay una encarnizada controversia en torno al diagnóstico, que va del fraude categórico a la aceptación de unos cuantos casos excepcionales y genuinos rodeados de aspirantes tutifruti, pasando por *folie à deux* (en la que un psiquiatra ingenuo alienta sin querer a un paciente trastornado a cultivar los síntomas). Tras varios años de estudiar el fenómeno (y a la gente que estudia y trata el fenómeno), el psicólogo Nicholas Humphrey y yo decidimos que todos tienen razón.[36] Hay fraudes, exageraciones, terapeutas crédulos con pacientes ansiosos, y sí, unos cuantos casos en los que la enfermedad parece haber existido al menos de forma rudimentaria antes de cualquier interpretación de parte de un interlocutor fascinado. Esto no es sorprendente si reconocemos que es la intensificación de algo muy normal que todos experimentamos en diferente grado. La mayoría de nosotros llevamos vidas bastante diferentes: en el trabajo, en la casa, en el juego, y en cada contexto adquirimos hábitos y recuerdos que a veces no viajan bien a los otros contextos.

Como el sociólogo Erving Goffman narra en su libro clásico sobre el sujeto, *La presentación de la persona en la vida cotidiana,* todos nos presentamos como personajes en dramas de la vida real (profesor Dennett, Dan que vive en la acera de enfrente, papá, abuelo, etc.) y sin esfuerzo reclutamos a los actores de reparto, que también se presentan a sí mismos.[37] Nos damos unos a otros argumentos con los que nos podemos desenvolver sin reparos y nos hacemos cómplices en nuestras mutuas campañas de presentación del yo, o podemos perturbar estos escenarios que avanzan con soltura al actuar fuera de personaje, con resultados incómodos, cómicos o peor. Se necesitan nervios de acero para eso. ¿Te imaginas pedirle a alguien a quien te acaban de presentar en una fiesta que te muestre una identificación, la licencia de conducir o el pasaporte? O bien, en la dirección opuesta, ¿te imaginas abrazarla apasionadamente? Cuando se coloca a la gente en situaciones extremadamente difíciles, a veces adopta

medidas extremas, y lo que empieza como un fingimiento desesperado se convierte en algo que hace casi automáticamente. Cuando las cosas se ponen difíciles, simplemente *te vas,* y dejas atrás un distinto centro de gravedad narrativa, un diferente personaje, mejor equipado para ocuparse del problema.

Humphrey y yo nos dimos cuenta de que cuando entrevistas a alguien que supuestamente sufre del trastorno de identidad disociativo, debes esperar encontrar a un ingenuo que domina el arte de desviar las indagaciones molestas. Para hacerle las preguntas que piden respuestas a gritos[38] o —mejor— poner pequeñas trampas para ver si de verdad un álter no tiene recuerdos de lo que el otro álter estaba haciendo y diciendo, tienes que ser descaradamente grosero y arriesgarte a ofenderlo, así que con toda probabilidad te descubrirás a ti mismo siguiéndole el juego educadamente, convertido en un cómplice al que reclutó en un abrir y cerrar de ojos. Los estafadores lo hacen a propósito, con gran destreza; también las víctimas inocentes de este trastorno de la personalidad, pero ellas sin darse cuenta de lo que están haciendo. Igual todos los demás, hasta cierto punto. Pero *sus* álter son sólo personajes ficticios, ¿verdad? La persona *real* es el huésped, ¿verdad? Bueno, no es tan sencillo.

Esa cosa que es *tú,* la persona que interpretas, ya sea que interpretes múltiples papeles o sólo un papel monolítico, es tu centro de gravedad narrativa. Es como tus amigos te reconocen («¡Hoy estás fuera de ti!») y como te ves tú en general, pero también está un poco idealizado («¡Santo Dios! ¿*Yo* hice eso? ¡Yo nunca haría algo así!»). Los novelistas profesionales, como los estafadores, crean narrativas con una ingeniosa y deliberada atención a los detalles. Todos los demás somos unos aficionados talentosos, que hilamos nuestros relatos hábilmente pero por lo general sin darnos cuenta, un poco como una araña teje su tela. Es naturaleza, no arte. No es tanto que *nosotros,* usando nuestros cerebros, hilemos nuestras historias, como que nuestros cerebros, usando historias, nos hilan *a nosotros.* Hay un núcleo de biografía verdadera indudable, seguro que sí, pero con los años grandes partes de ella pueden darse por perdidas, se vuelven inertes y dejan de tener la menor importancia para quien eres ahora. Puede ser que activamente reniegues de una parte, la eches por la borda, la «olvides», en el proceso de mantenerte y superarte.[39]

Piensa cuán fácilmente podrías responder algunas preguntas sobre tu pasado. ¿Alguna vez has bailado con una estrella de cine? ¿Alguna vez has ido a París? ¿Alguna vez has paseado en un camello? ¿Alguna vez has matado a alguien estrangulándolo con las puras manos? Éstas son fáciles casi para todos, sea sí o no la respuesta. (Imagina a alguien a quien le hacen esta última pregunta y que se queda pensativo rascándose la barba antes de responder. ¡Aléjate de él!) Conocemos las respuestas a estas preguntas porque sabemos suficiente de nosotros mismos —de nuestros *yos*— para saber que si alguna vez hubiéramos bailado con una estrella de cine, ido a París, paseado en un camello o

estrangulado a alguien, ahora lo recordaríamos. Cuando «no viene nada a la mente» interpretamos esta ausencia como una negativa. (¿De qué otro modo podríamos estar tan seguros? ¿Llevas una lista de todas las cosas que nunca has hecho? ¿Todos los lugares a los que nunca has ido?) Contrasta esas preguntas con estas otras que sólo superficialmente se parecen: ¿Alguna vez has bailado con alguien que se apellide Gómez? ¿Alguna vez has ido a una miscelánea que vendiera abrillantador de suelos? ¿Alguna vez has paseado en un Chevrolet azul? ¿Alguna vez has roto una taza de café blanca? Puede ser que algunas sean fáciles de responder, otras te lleven a decir que no tienes ni idea, y otras pueden moverte a dar respuestas falsas sin que siquiera te des cuenta, de tan poco importantes que son. ¿Por qué tendrías que recordar si hiciste cualquiera de estas cosas? Mucho de lo que nos ha pasado sencillamente no es memorable; nuestros centros de gravedad narrativa se han deshecho de muchas cosas que hemos hecho, buenas y malas. También, de manera del todo inocente se han añadido muchas cosas que nunca nos pasaron, porque por una u otra razón nos ha convenido. Lo que eres es esa suma rodante de experiencia y talento, intención solemne y fantasía ensoñadora, unidos en un cerebro y un cuerpo y a los que se llama con el nombre que te pusieron. La idea de que, además, hay una pepita de *ti,* o un ego, espíritu o alma es una fantasía atractiva, pero nada que se necesite para entender a la gente, sus sueños y esperanzas, su heroísmo y sus pecados.

El centro de gravedad narrativo puede no ser una misteriosa pepita de cosas mentales, pero si sólo es una abstracción, ¿puede estudiarse científicamente? Sí, se puede.

63. Heterofenomenología

La heterofenomenología no es una bomba de intuición, sino otro ejemplo de *montaje* que vale mucho la pena poner en su lugar antes de abordar algunas preguntas difíciles. El estudio de la conciencia humana implica fenómenos que a primera vista parecen ocurrir en algo así como otra dimensión: la privada y subjetiva dimensión de «primera persona» que cada uno de nosotros ocupa con respecto a nuestra propia conciencia, y a la que nadie más puede tener acceso directo. ¿Entonces cuál es la relación entre las metodologías objetivas, de «tercera persona», estándar para estudiar a los meteoritos o los imanes (o bien el metabolismo o la densidad ósea de los humanos), y las metodologías para estudiar la conciencia humana? ¿Tenemos que crear alguna ciencia alternativa radical o revolucionaria, o pueden los métodos estándar extenderse de tal modo que se haga justicia a los fenómenos de la conciencia humana? Defiendo la afirmación de que hay una sencilla extensión cautelosa de la ciencia objetiva que generosamente abarca *todo* el campo de la conciencia humana y hace justicia a todos los datos sin tener que abandonar jamás las reglas y las limitaciones de los métodos experimentales que han funcionado tan bien en el resto de la ciencia. Esta metodología de tercera persona, la heterofenomenología (fenomenología de un *otro,* no de uno mismo) es la manera razonable de tomar el punto de vista de la *primera* persona con tanta seriedad como legítimamente se puede.

¿Por qué el nombre multisilábico? «Fenomenología» originalmente significaba un catálogo de fenómenos, de una u otra clase, antes de que hubiera una buena teoría para ellos. En el siglo XVI, William Gilbert recopiló una buena fenomenología del magnetismo, pero fue siglos antes de que todos los fenómenos magnéticos que describió con esmero pudieran explicarse. A principios del siglo XX, Edmund Husserl y un grupo de psicólogos y filósofos sobre los que tenía influencia adoptaron el término «Fenomenología» (con «F» mayúscula) para referirse a un supuesto estudio científico de los fenómenos de la experiencia subjetiva, que debían observarse mediante un método introspectivo de «primera persona» que trataba de ser neutral, no depender de ninguna teoría y no tener muchas suposiciones. La escuela de pensamiento permanece hasta nuestros días, casi siempre criticada o ignorada, por buenas y malas razones. A pesar de algunos resultados muy seductores que bien merecen una mayor exploración, por ser un enfoque de primera persona ha sido rechazada por la ciencia empírica, objetiva, que insiste en contar con datos accesibles a todos los investigadores. Sin embargo, podemos estudiar la conciencia objetivamente, y el método no es más que una sencilla vuelta de tuerca a la fenomenología, así que lo llamo heterofenomenología, para contrastarlo con la *autofenomenología* husserliana. La heterofenomenología es el estudio de los fenómenos de primera persona desde el punto de vista de tercera persona que emplea la ciencia objetiva.

Evidentemente la diferencia clave entre los experimentos con rocas, rosas y ratas, por un lado, y los experimentos con sujetos humanos despiertos y dispuestos a cooperar, por el otro, es que estos últimos pueden comunicarse mediante el lenguaje y por lo tanto pueden colaborar con los experimentadores haciendo sugerencias, interactuando verbalmente y diciéndoles qué sienten bajo diversas condiciones controladas. Ése es el meollo de la heterofenomenología: aprovecha nuestra capacidad de *realizar* e *interpretar* actos de habla, con lo que arroja un catálogo de *lo que el sujeto cree verdadero sobre su experiencia consciente.* Este catálogo de creencias le da cuerpo al *mundo heterofenomenológico* del sujeto, el mundo según S, el mundo subjetivo de un sujeto. El conjunto total de los detalles de la heterofenomenología, junto con todos los datos que podemos obtener sobre los sucesos concurrentes en los cerebros de los sujetos y en el medio que los rodea, comprenden el conjunto completo de datos que una teoría de la conciencia humana debe explicar. No deja fuera ningún fenómeno objetivo ni ningún fenómeno subjetivo de conciencia.

La interpretación que se requiere para convertir datos *crudos* sobre los sonidos del habla y pulsaciones de botones en informes y expresiones de creencia implica adoptar la **postura intencional**: requiere la hipótesis provisional de que el sujeto es un agente cuyas acciones están guiadas racionalmente por creencias y deseos que también son racionales, dada la historia perceptual y las necesidades del sujeto. Por ejemplo, las limitaciones de la postura intencional pueden percibirse claramente en las precauciones que habitualmente se toman en esos experimentos para impedir que los sujetos tengan experiencias que pudieran darles o creencias o deseos que tendieran a inclinar sus respuestas de maneras que distorsionarían nuestra interpretación de sus acciones: les ocultamos qué esperamos que digan, por ejemplo, y al mismo tiempo tomamos medidas para asegurarnos de que entienden las tareas que les asignamos. Esta adopción de la postura intencional no es un asunto irremediablemente subjetivo y relativista. Se pueden articular reglas de interpretación; se pueden fijar y cumplir normas de acuerdo intersubjetivo sobre la interpretación; pueden identificarse desviaciones; la inevitable suposición de racionalidad puede expresarse cautelosamente y ser tratada como una suposición explicable desde el punto de vista de la teoría evolutiva, ajustable y defendible.[40]

Esto no es una propuesta de una nueva metodología para estudiar la conciencia. Simplemente soy consciente de los métodos que los investigadores en psicología cognitiva, psicofísica (que estudia las relaciones entre estímulos físicos y reacciones subjetivas) y neurociencia suelen adoptar, y los explico y defiendo. Estos métodos, si se comprenden y se siguen correctamente, eluden la necesidad de una ciencia de la conciencia de «primera persona» radical o revolucionaria, y no dejan fenómenos residuales de conciencia que sean inaccesibles a un estudio científico controlado.

¿Con qué clase de cosas nos compromete esta metodología? Más allá de las cosas no problemáticas con las que toda la ciencia está comprometida (neutrones y electrones, relojes y microscopios), nos compromete sólo con *creencias* —las creencias que expresan los sujetos y que se consideran constitutivas de su subjetividad— y con *deseos* —los deseos de cooperar con los experimentadores y de decirles la verdad con tanta franqueza como sea posible (una parte importante de este método supone controlar estas creencias y deseos; debe desecharse cualquier resultado experimental que dé muestras de que esto no se ha conseguido). ¿Qué clase de cosas son las creencias y los deseos? Podemos postergar al máximo el momento de definirnos al respecto —en espera de que se confirme la teoría— si tratamos las creencias y sus contenidos u objetos como *ficciones de teórico* o *abstracciones* parecidas a los centros de masa, el ecuador y los paralelogramos de fuerzas.

Los avistamientos de sirenas son sucesos reales, por mal que se hayan descrito, mientras que las sirenas no existen. Del mismo modo, un catálogo de creencias acerca de la experiencia no es lo mismo que un catálogo de las experiencias mismas. El filósofo Joseph Levine ha objetado que «las experiencias conscientes mismas, no simplemente nuestros juicios verbales acerca de ellas, son los datos primarios que una teoría debe responder».[41] Esto no puede estar bien. ¿Cómo podríamos catalogar las experiencias mismas antes de tener una teoría? Piensa en los indicios que obtenemos cuando ponemos a sujetos en situaciones experimentales y los interrogamos (y les pedimos que realicen cualesquiera otras acciones que queramos). Estas fuentes encajan unas con otras naturalmente por la cascada de interpretaciones que tenemos que realizar, ordenadas aquí de la menos cruda a la más cruda:

a) «las experiencias conscientes mismas»,
b) las creencias acerca de estas experiencias conscientes,
c) los «juicios verbales» que menciona Levine,
d) las emisiones de una u otra clase que (pueda interpretarse que) expresan esos juicios verbales.

En un sentido de «primario», las emisiones son datos primarios —sonidos y movimientos registrados—. Las lecturas del electroencefalograma y las lecturas de las imágenes obtenidas por resonancia magnética funcional y demás pueden añadirse a los datos primarios, si las dircunstancias lo permiten. Podemos llegar a *c* y *b* mediante métodos de interpretación confiables, así que tenemos un catálogo de creencias de sujetos acerca de cómo es ser ellos bajo estas condiciones. Sin embargo, ¿deberíamos seguir adelante con *a* antes de tener una teoría? No es buena idea, por dos razones.

En primer lugar, si *a* supera a *b* —si tienes experiencias conscientes que no crees tener—, esas experiencias conscientes extras son tan inaccesibles *para ti* como para los observadores externos.

Así, la alternativa que Levine proponía no recoge más datos provechosos que la heterofenomenología.

En segundo lugar, si *b* supera a *a* —si crees tener experiencias conscientes que en realidad no tienes—, entonces serán tus creencias lo que tendremos que explicar, no tus inexistentes experiencias.

Entonces, si nos atenemos a los principios heterofenomenológicos y tratamos *b* como el máximo conjunto de datos *primarios,* evitaremos cualquier compromiso con datos espurios, y al mismo tiempo garantizaremos que se incluyan todos los fenómenos accesibles a *cualquiera.*

¿Y si algunas creencias no pueden expresarse en juicios verbales? No hay nada que impida a los heterofenomenólogos y a los sujetos inventar juntos modos análogos o no lingüísticos de expresar creencias. Por ejemplo:

Traza una raya vertical a través del segmento de la siguiente línea que indique qué tan intensa [en una u otra dimensión] es la experiencia:

Apenas perceptible ——————————————————— Abrumadora

O los sujetos pueden apretar un botón con presión variable para indicar la intensidad de un dolor (o de la ansiedad, del aburrimiento o incluso de desconfianza en el experimento). Y luego hay que medir una gran cantidad de variables fisiológicamente dependientes, desde la respuesta galvánica de la piel y la frecuencia cardiaca hasta cambios en la expresión facial y en la postura. Y si tú, el sujeto, crees que *todavía* hay residuos inefables que no se han expresado tras haber agotado estos métodos, puedes decírselo a los heterofenomenólogos, que pueden agregar esa creencia a la lista de creencias en tus datos primarios:

S dice tener creencias inefables acerca de X.

Si esta creencia es verdadera, la ciencia tendrá la obligación de explicar qué son tales creencias y por qué son inefables. Si esta creencia es falsa, de todas maneras la ciencia tendrá que explicar por qué S cree (falsamente) que existen estas creencias inefables

específicas.[42]

64. Mary, la científica del color: una muleta de razonamiento al descubierto

El experimento mental del filósofo australiano Frank Jackson sobre Mary, la científica del color, a menudo llamado «el argumento del conocimiento», ha estado bombeando las intuiciones de los filósofos con notable vigor desde que apareció en 1982. Tan sólo por volumen y fiabilidad, debería contar como una de las bombas de intuición más exitosas jamás inventadas por un filósofo analítico. Es un clásico, siempre presente en la lista de lecturas obligatorias de cursos de licenciatura sobre filosofía de la mente en todo el mundo de habla inglesa, y se han publicado varias antologías de muchas páginas con ensayos que reflexionan sobre sus implicaciones. Es interesante señalar que su autor después se retractó y ha declarado que ya no acepta su conclusión, pero eso no le ha restado popularidad.

Hela aquí en su totalidad; tal vez está sacada un poco de contexto, pero no por eso deja de ser clarísima:

> Mary es una brillante científica que, por alguna razón, está obligada a investigar el mundo desde una habitación blanca y negra *a través* de un monitor de televisión en blanco y negro. Se especializa en la neurofisiología de la visión y adquiere, supongamos, toda la información física que puede obtenerse sobre lo que pasa cuando vemos unos jitomates maduros o el cielo y usamos términos como «rojo», «azul», etc. Descubre, por ejemplo, específicamente qué combinaciones de longitudes de onda procedentes del cielo estimulan la retina, y exactamente cómo esto produce *a través* del sistema nervioso central la contracción de las cuerdas vocales y la expulsión de aire de los pulmones que se traduce en la pronunciación del enunciado «El cielo es azul». (Difícilmente podrá negarse que en principio es posible obtener toda esta información física de una televisión en blanco y negro; de otro modo la universidad abierta habría tenido que usar, *necesariamente,* una televisión a color.) ¿Qué va a pasar cuando se libere a Mary de su habitación blanca y negra o se le dé un monitor de televisión a color? ¿*Aprenderá* algo o no? Parece de lo más obvio que aprenderá algo acerca del mundo y nuestra experiencia visual de él. Pero entonces sería inevitable concluir que su conocimiento anterior estaba incompleto. Sin embargo, ella tenía *toda* la información física. *Ergo,* hay algo más que eso, y el fisicalismo [*i. e.,* el materialismo, la negación del dualismo] es falso.[43]

¿Es una buena bomba de intuición? Giremos todas las perillas y veamos qué la hace funcionar. De hecho eso nos tomaría mucho tiempo, pero el trabajo ya se ha hecho en la considerable bibliografía sobre el tema. Aquí sencillamente ilustraré unas cuantas perillas

que necesitan examinarse y te dejaré de ejercicio obtener el resultado. (Si quieres, puedes contrastar tus resultados con la bibliografía publicada; en la sección de fuentes documentales se mencionan dos antologías recientes. ¿Se te ocurre algún nuevo giro?) Hace más de 20 años llevé a cabo una exploración preliminar de las perillas y emití un veredicto aguafiestas que en gran parte se ha ignorado o despreciado: «Como buen experimento mental, su conclusión es obvia de inmediato incluso para los no iniciados. De hecho es un mal experimento mental, ¡una bomba de intuición que en realidad nos alienta a malinterpretar sus premisas!».[44] Veamos si es cierto. Sostengo que imaginar correctamente el escenario es mucho más difícil de lo que la gente supone, de modo que imaginan algo más fácil y a partir de esa base equivocada extraen sus conclusiones.

> Primera perilla: «una habitación blanca y negra *a través* de un monitor de televisión en blanco y negro».

Debemos suponer que lleva guantes negros o blancos y que se le ha prohibido verse mientras se baña, pero en todo caso la idea de cortar todas las «fuentes externas» de color es triste. ¿Tendríamos que montar algún dispositivo que impidiera que se frotara los ojos (para crear fosfenos —inténtalo—)? ¿Y no podría haber tenido colores en sus sueños antes de haber visto en realidad los colores? Si no, ¿por qué no? ¿Los colores tienen que «entrar a través de los ojos» antes de que pueda «almacenarlos» en su cerebro? Detrás de esta simple sugerencia hay una maraña de mala teoría vulgar del color.

> Segunda perilla: «adquiere, supongamos, toda la información física que puede obtenerse sobre lo que pasa cuando vemos unos jitomates maduros o el cielo y usamos términos como "rojo", "azul", etcétera».

¿*Toda* la información física que puede obtenerse? ¿Qué tanto es eso? ¿Es como tener todo el dinero del mundo? ¿Cómo sería? No es fácil imaginarlo, y nada que sea menor a todo servirá para que el pensamiento mental consiga lo que se propone. Debe incluir toda la información acerca de toda la variación de respuestas en todos los cerebros, entre ellos el suyo, e incluir especialmente todas las reacciones emocionales o afectivas a todos los colores en todas las condiciones. Así sabrá con magnífico detalle qué colores la calman, cuáles le molestan, cuáles le llegarían a gustar si los viera, cuáles la distraen, cuales la repelen, etc. ¿Tiene prohibido realizar experimentos consigo misma (sin hacer trampa, sin meter a escondidas ninguna cosa de color a su celda)? Si no te imaginaste todo esto (y más), no seguiste las instrucciones. Es como si te piden que concibas un quiliógono y en vez de eso te imaginas un círculo. De uno de estos ejercicios de representación mental se

siguen muchas implicaciones que no se siguen del otro. En este ejemplo, pongamos por caso, ¿se espera que ignoremos el hecho de que si Mary adquiriera toda esta información se encontraría con toda seguridad en un estado absoluto de colapso mental, agobiada con miles de artículos y diagramas enciclopédicos?

Si Jackson hubiera estipulado que Mary tenía la propiedad divina de ser «físicamente omnisciente» —no sólo sobre los colores sino sobre todos los hechos físicos en todos los niveles, del quark a la galaxia—, muchos lectores, si no es que todos, lo rechazarían y dirían que imaginar esa hazaña es demasiado fantástico para tomarlo en serio. Sin embargo, estipular que Mary conoce todos los datos *solamente* físicos *acerca de la visión del color* no es mucho menos fantástico.

—Imagina que Mary tiene un billón de cabezas…
—¡No seas tonto!
—Bueno, que sea un millón.
—De acuerdo.
(¿De veras?)

En un intento previo de dramatizar este problema de imaginación, animé a la gente a pensar en un final divergente:

> Y así, un buen día, los secuestradores de Mary decidieron que había llegado el momento de que pudiera ver los colores. Para engañarla, prepararon un plátano de color azul brillante para presentárselo como su primera experiencia cromática en la vida. Mary lo miró y dijo: «¡Hey! ¡Me han querido engañar! Los plátanos son amarillos, ¡pero éste es azul!» Los secuestradores se quedaron boquiabiertos. ¿Cómo lo hizo? «Fácil», respondió ella, «no deben olvidar que yo sé *todo* —absolutamente todo— lo que jamás pueda saberse sobre las causas y efectos físicos de la visión en color. Así que, por supuesto, antes de que me trajeran el plátano yo ya tenía anotado, con todo lujo de detalles, exactamente qué impresión física produciría un objeto amarillo o un objeto azul (o un objeto verde, etc.) en mi sistema nervioso. Así que yo ya conocía de antemano qué *pensamientos* tendría (porque, después de todo, la "mera disposición" a pensar sobre esto o aquello no es uno de sus famosos qualia, ¿o sí?). Mi experiencia del azul no me sorprendió en lo más mínimo (lo que me sorprendió fue que quisieran gastarme una broma tan barata). Me doy cuenta de que *les resulta muy difícil imaginar* que yo pudiera saber tanto sobre mis disposiciones reactivas que el modo en que me afectó el azul no me sorprendiera. Claro que les resulta difícil. ¡Para cualquiera es difícil imaginar las consecuencias de que alguien tenga absolutamente todo el conocimiento físico acerca de lo que sea!»[45]

Por lo general se acepta que las cosas no podrían marchar así. Jackson lo expresó de un modo que desarma: «Parece de lo más obvio que aprenderá algo acerca del mundo y nuestra experiencia visual de él». O como dicen George Graham y Terry Horgan: «Accedemos: no cabe duda [¡ring!] de que debería estar sorprendida y encantada».[46] Es un error, y eso es lo que pasa con Mary como experimento mental. Sencillamente se siente tan bien concluir que Mary tiene una revelación de *alguna* clase la primera vez que ve colores que nadie quiere tomarse la molestia de mostrar que así es como debe seguir la historia. De hecho, no es necesario que vaya por ahí en absoluto.

La bomba de intuición de Jackson deja magníficamente al descubierto un montón de pensamiento ingenuo acerca de la naturaleza de la experiencia de color y el cerebro; seguramente eso le es muy útil a la gente la mayor parte del tiempo, así que podemos conceder que extrae muy bien algunas de las implicaciones de la teoría vulgar. Sin embargo, su objetivo era refutar una hipótesis acerca de la capacidad de las ciencias físicas para explicar todos los fenómenos del color. Desde luego, en cualquier situación de la vida real, alguien en la posición imaginaria de Mary aprendería algo nuevo porque, por mucho que supiera del color, habría muchísimos hechos acerca de los efectos físicos del color que no sabría. Es sólo por el estipulado caso extremo que el «de lo más obvio» de Jackson y el «no cabe duda» de Graham y Horgan están tan fuera de lugar. Si sigues inclinado a pensar que mi sugerencia de final alternativo para el relato tendría que ser imposible, ve si tú puedes encontrar una razón argumentada para tu creencia. Sería interesante ver si se te ocurre alguna consideración que se les haya escapado a los cientos de filósofos que durante años han trabajado incansablemente en esto (desde luego, ese mismo hecho podría servir para mostrar que ésta es a fin de cuentas una bomba de intuición maravillosa: ha dado empleo a los filósofos a lo largo de tres décadas).

Resumen

Un problema acuciante para el estudio científico de la conciencia ha sido el hecho de que ¡*todo el mundo* es un experto! En realidad no, desde luego, pero prácticamente cualquiera que haya reflexionado sobre el tema más de tres minutos parece pensar que los resultados de esas reflexiones son tan fidedignos como los de cualquier experimento de alta tecnología o cualquier montaña de estadísticas. Puede ser realmente cómico oírlos durante las sesiones de preguntas y respuestas en las conferencias científicas *corrigiendo* con firmeza los *errores* que acaban de detectar en el trabajo del presentador citando una experiencia reciente que ellos (piensan que) tuvieron. Si, como estos individuos normalmente creen, fuéramos jueces infalibles de la naturaleza de nuestras propias experiencias personales, ¡tendrían razón!

Sin embargo, uno *puede* recordar a medias, malinterpretar, describir equivocadamente sus experiencias más íntimas, secretamente llevado por alguna pizca de teoría convincente pero poco fidedigna. Presento una simple demostración para realizar en casa que te puede sorprender. Siéntate frente a un espejo de modo que puedas vigilar que cumples las instrucciones, a saber, quedarte mirando tus propios ojos, fijando en ellos la atención como si se tratara de un blanco, en vez de dejar que se desvíen porque los atraiga algo que ocurra alrededor. Ahora, sin ver, toma una carta de alguna baraja bien mezclada y sostenla con el brazo estirado, *con la cara hacia ti,* justo afuera de los límites de tu visión periférica. Agita la carta. Sabrás que lo estás haciendo, pero no vas a verlo, por supuesto. Empieza a mover la carta hacia tu campo de visión, agitándola mientras lo haces. Primero puedes ver movimiento (la agitación), ¡*pero nada de color!* No puedes saber si es una carta roja o negra o si tiene una figura, y ciertamente no puedes identificar su número. Conforme la mueves más y más hacia el centro, te sorprenderá ver que debe estar casi frente a ti para que puedas identificar su color, o si tiene o no una figura. Mientras más y más se acerca al punto en el que has fijado tu atención, debes concentrarte en no hacer trampa y mirarla de soslayo. Cuando finalmente puedes identificarla, está justo frente a ti, prácticamente. ¿Te sorprende? No conozco a nadie que no se haya sorprendido la primera vez que experimentó esto. Toda la vida has pensado que tu visión era en líneas generales igual de detallada y colorida «en todo tu campo de visión central», y ahora te enteras de que aunque «parece lógico» y parece estar confirmado por la introspección causal, simplemente no es cierto. Ésta es sólo una dimensión, un fenómeno entre muchos, en la que la aparentemente intensa, continua y detallada presentación del mundo en (nuestra) conciencia visual es una ilusión. La moraleja es inequívoca: no pienses que entiendes los fenómenos de la conciencia hasta que veas lo que la ciencia ha descubierto sobre ella en los últimos años. Las teorías de café de los filósofos que ignoran esta moraleja son en el mejor de los casos insignificantes, pero casi siempre enredadas y confusas. Lo que «aprendes» sobre tu

conciencia «mediante la introspección» es una porción menor pero poderosamente engañosa de lo que *nosotros* podemos aprender sobre *tu* conciencia si adoptamos el marco heterofenomenológico y estudiamos sistemáticamente la conciencia.

Todavía hay un sinfín de dificultades en busca de solución, problemas difíciles que no son el Problema Difícil. Si cuando hayamos resuelto todos estos problemas «fáciles» todavía hay un residuo profundamente misterioso, será momento de reconsiderar nuestro punto de partida y tratar de encontrar alguna desviación radical de las suposiciones vigentes sobre la biología, la física y hasta la lógica. En el ínterin, veamos qué tan lejos podemos llegar con la ciencia de siempre, la ciencia que nos ha traído nuestra actual comprensión de todo, desde los asteroides y las placas tectónicas hasta la reproducción, el crecimiento, la regeneración y el metabolismo de las cosas vivientes.

[1] R. Jackendoff, *Consciousness and the Computational Mind,* The MIT Press, Cambridge, 1987, y D. C. Dennett, *Consciousness Explained, op. cit.*

[2] C. Dennett, *Consciousness Explained, op. cit.*

[3] Para los argumentos, véanse D. C. Dennett, *Consciousness Explained, op. cit.,* y *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness,* The MIT Press, Cambridge, 2005.

[4] Alguien que leyó uno de los primeros borradores de este libro pensó que todo esto era una invención mía. Si compartes esa sospecha, revisa el largo y aburridamente serio artículo sobre los zombis filosóficos en la augusta *Stanford Encyclopedia of Philosophy,* que puede consultarse en internet: <http://plato.stanford.edu/entries/zombies/>.

[5] G. W. F. von Leibniz, «Monadology», en *The Monadology and Other Philosophical Writings,* Oxford University Press, Oxford, 1898. Las cursivas son mías.

[6] Amber Ross, *Inconceivable Minds,* tesis de doctorado en Filosofía, University of North Carolina, Chapel Hill, 2013.

[7] Ésta es otra de esas situaciones en que a algunos filósofos les gustaría acusarme de confundir una pregunta metafísica con una epistemológica: «No preguntes *cómo sabríamos* que alguien es un zombi: ¡pregunta *qué es* ser un zombi!» La metafísica, aislada de ese modo de la epistemología, es en el mejor de los casos un ejercicio de fantasía.

[8] D. C. Dennett, *Consciousness Explained, op. cit.*

[9] T. Nagel, «What is It Like to Be a Bat?», *Philosophical Review,* 83 (4): 435-450, 1974.

[10] Véase D. C. Dennett, «Evolution, Error and Intentionality», en D. Partridge y Y. Wilks (comps.), *Sourcebook on the Foundations of Artificial Intelligence,* University of New Mexico Press, Albuquerque, 1988, pp. 190-211.

[11] N. Block, «Troubles with Functionalism», en W. Savage (comp.), *Perception and Cognition: Issues in the Foundations of Psychology,* vol. 9, University of Minnesota Press, Minneapolis, 1978, p. 281.

[12] T. Nagel, «What Is It Like to Be a Bat?», *op. cit.*

[13] Una exploración pionera de este fenómeno puede encontrarse en S. Dehaene y J. F. Marques, «Cognitive Euroscience: Scalar Variability in Price Estimation and the Cognitive Consequences of Switching to the Euro», *Quarterly Journal of Experimental Psychology,* 55: 705-731, 2002.

[14] G. Graham y T. Horgan, «Mary Mary Quite Contrary», *Philosophical Studies,* 99 (1): 59-87, 2000, p. 73.

[15] Véase D. Chalmers, «Facing Up to the Problem of Consciousness», *Journal of Consciousness Studies,* 2 (3): 200-219, 1995.

[16] Para una respuesta del tamaño de un libro a esta pregunta, véase M. Hurley, D. C. Dennett y R. B. Adams Jr., *Inside Jokes: Using Humor to Reverse-Engineer the Mind,* The MIT Press, Cambridge, 2011.

[17] R. Wright, *Nonzero: The Logic of Human Destiny,* Pantheon, Nueva York, 2000, p. 398.

[18] John Northern Hilliard, *Card Magic,* Carl W. Jones, Minneapolis, 1938.

[19] J. Searle, «Minds, Brains and Programs», *op. cit.,* p. 417.

[20] J. Searle, «Turing the Chinese Room», en T. Singh (comp.), *Synthesis of Science and Religion, Critical Essays and Dialogues,* Bhaktivedanta Institute, San Francisco, 1988, p. 136.

[21] D. C. Dennett, «The Milk of Human Intentionality», *op. cit.*

[22] D. Hofstadter y D. C. Dennett (comps.), *The Mind's I. Fantasies and Reflections on Self and Soul,* Basic Books, Nueva York, 1981.

[23] Doug apuntó directamente a la frase «trozos de papel» en el ensayo de Searle, y mostró cómo alentaba a la gente a subestimar en muchos órdenes de magnitud el tamaño y la complejidad del *software* implicado. Esta

crítica ocupaba un lugar importante en su comentario sobre Searle en nuestro libro, y dio lugar a una feroz respuesta de Searle en las páginas del *New York Review of Books,* porque, aunque habíamos reproducido correctamente su artículo, a Doug se le escapó un error en su comentario y puso «unos cuantos papelitos» donde Searle había dicho «unos cuantos trozos de papel», y esto, sostenía Searle, ¡tergiversaba completamente su argumento! Si Searle estaba en lo cierto, si ese pequeño cambio involuntario transformaba la maquinaria, esto en realidad demostraba, de alguna manera, lo que nosotros sosteníamos: si un ajuste tan pequeño inutiliza o posibilita un experimento mental, debería señalárseles a todos aquellos cuyas intuiciones se van a bombear. Véase J. Searle, «The Myth of the Computer», *The New York Review of Books,* 29 (7): 3-6, 1982.

24 Lee la demostración de Hofstadter, propia de un virtuoso, en D. Hofstadter y D. C. Dennett (comps.), *The Mind's I, op. cit.,* p. 460, donde prueba que la habitación china es una bomba de intuición con (por lo menos) cinco perillas, que procede a girar, una por una, para obtener una familia de distintos experimentos mentales con las distintas conclusiones a las que se prestan. En las reflexiones de cierre observé: «Cada posición de los diales de nuestra bomba de intuición da lugar a una narrativa ligeramente distinta, con distintos problemas perdiéndose a la distancia y diferentes moralejas. En qué versión o versiones confiar es algo que deberá decidirse examinándolas cuidadosamente para ver qué rasgos de la narrativa están haciendo el trabajo. Si las simplificaciones exageradas son la *fuente* de las intuiciones, y no sólo recursos para eliminar complicaciones irrelevantes, deberíamos desconfiar de las conclusiones que se nos invita a sacar».

25 M. Gell-Man, *The Quark and the Jaguar: Adventures in the Simple and the Complex,* St. Martin's, Nueva York, 1995.

26 Como varias veces se ha señalado, esta convicción se hace eco de la de Descartes, quien en su *Discurso del método,* allá por 1637, propuso que la mejor manera de distinguir entre una máquina y una persona con un alma inmaterial era tener una conversación con ella. «Se puede concebir que pudiera hacerse una máquina de tal manera que profiera palabras, y aun que pronuncie algunas con ocasión de las acciones corporales que causan algún cambio en sus órganos —como, por ejemplo, si se la toca en una parte que pregunte lo que se quiere decirle, y si en otra, que grite que se le hace daño y otras cosas semejantes—, pero no que arregle las palabras de diversos modos para responder según el sentido de cuanto en su presencia se diga, como pueden hacer aun los más estúpidos de los hombres.»

Nadie sabe si para su bomba de intuición Turing se inspiró en la de Descartes.

27 J. Searle, «Minds, Brains and Programs», *op. cit.,* p. 419.

28 *Idem.*

29 D. C. Dennett, «Turing's "Strange Inversion of Reasoning"», en B. Cooper (comp.), *Alan Turing—His Work and Impact,* Elsevier, Nueva York, 2013, pp. 569-573.

30 J. Searle, «Minds, Brains and Programs», *op. cit.,* p. 418.

31 *Idem.*

32 Agradezco al lingüista, sinólogo y erudito David Moser por este ejemplo, que se creó en China en el momento en que yo estaba tratando de encontrar algo justamente así.

33 D. Hume, *A Treatise of Human Nature,* ed. de L. A. Selby-Bigge, Clarendon, Oxford, 1964, I, IV, secc. 6.

34 «I'd rather have a free bottle in front of me than a prefrontal lobotomy», quiasmo que juega más con la posición de las sílabas y el sonido de las palabras que con su significado, así que el chiste inevitablemente se pierde en español. [T.]

35 D. Lewis, «Truth in Fiction», *American Philosophical Quarterly,* 15 (1): 37-46, 1978.

36 N. Humphrey y D. C. Dennett, «Speaking for Ourselves: An Assessment of Multiple Personality Disorder», *Raritan: A Quarterly Review,* 9 (1): 68-98, 1989.

37 E. Goffman, *The Presentation of Self in Everyday Life,* University of Edinburgh-Social Sciences Research Centre, Edimburgo, 1956.

38 Como éstas: ¿cómo deciden entre ustedes quién es el huésped y quiénes son álter? ¿Cómo te has explicado

a ti mismo que tu recuerdo de experiencias recientes consista de breves episodios de *estar ahí* rodeados de horas o días de olvido? ¿No te parece alarmante?

39 En el drama poético de T. S. Eliot *Asesinato en la catedral,* le preguntan a Becket sobre un suceso de su pasado y responde: «You talk of seasons that are past. I remember / Not worth forgetting» [Hablas de estaciones que han pasado. Recuerdo / que no eran para olvidarse].

40 D. C. Dennett, *Consciousness Explained, op. cit.*

41 J. Levine, «Out of the Closet: A Qualophile Confronts Qualophobia», *Philosophical Topics,* 22 (1-2): 107-126, 1994, p. 117.

42 La defensa de la heterofenomenología como la metodología preferida para el estudio científico de la conciencia ha dado lugar a una polémica considerable. Algunos investigadores la consideran una revolucionaria aclaración de las condiciones en que la ciencia puede estudiar la conciencia, algunos la consideran como un simple ensayo de lo evidente, y otros más siguen oponiéndose. Véase la sección de fuentes documentales, al final de este libro, para algo del mejor trabajo sobre el tema que haya aparecido hasta la fecha.

43 F. Jackson, «Epiphenomenal Qualia», *Philosophical Quarterly,* 32 (127): 127-136, 1982.

44 D. C. Dennett, *Consciousness Explained, op. cit.,* p. 398.

45 *Ibid.,* pp. 399-400.

46 G. Graham y T. Horgan, «Mary Mary Quite Contrary», *op. cit.,* p. 72.

VIII. Herramientas para pensar acerca del libre albedrío

Cuando más traicionero es el abismo entre la imagen manifiesta y la imagen científica es cuando se aborda el libre albedrío. Como con las preguntas de qué es el color, qué es *realmente,* y qué son *realmente* los dólares, cuando te pones a ver, la pregunta de si el libre albedrío es una ilusión o algo que en realidad tenemos nos invita a usar la imagen científica para investigar el tema, que está planteado en los términos tradicionales de la imagen manifiesta. Y la invitación ha sido aceptada con entusiasmo en los últimos años. Ha habido todo un coro de científicos eminentes diciendo a quemarropa que el libre albedrío es una ilusión: los neurocientíficos Wolf Singer, Chris Frith y Patrick Haggard; los psicólogos Paul Bloom y Daniel Wegner, y unos físicos más o menos respetados, Stephen Hawking y Albert Einstein. ¿Podrían estar equivocados tantos científicos brillantes? Muchos filósofos —no todos, y a lo mejor no la mayoría— dicen que sí. ¡Dicen que éste *es* un trabajo para la filosofía! ¿Tienen razón? Creo que sí.

Para variar, los científicos han estado cometiendo un error de principiante: confundir la imagen manifiesta con lo que podríamos llamar la *ideología vulgar* de la imagen manifiesta. La ideología vulgar del color es, hay que reconocerlo, un despropósito; el color no es lo que la mayoría de la gente piensa, pero eso no significa que el mundo manifiesto realmente no tenga colores; significa que los colores —los colores reales— son muy distintos de lo que la mayoría de la gente piensa que son. La ideología vulgar de la conciencia también es un despropósito: resueltamente dualista y misterial; si fuera *eso* lo que la conciencia tuviera que ser, entonces Wright tendría razón (véase el apartado 59): tendríamos que decir que la conciencia no existe. Sin embargo, no tenemos que tratar a la conciencia como «magia de verdad» —la que no existe, hecha de tejido maravilloso—; podemos admitir la realidad de la conciencia como un fenómeno si reconocemos que el vulgo todavía no tiene una ideología sensata acerca de ella. Del mismo modo, el libre albedrío no es lo que una parte de la *ideología vulgar* de la imagen manifiesta declara, una especie de aislamiento mágico con respecto a la causalidad. He comparado el libre albedrío en este sentido con la levitación, y uno de los defensores filosóficos de este despropósito de postura ha anunciado francamente que una libre elección es un «pequeño milagro». Estoy totalmente de acuerdo con el coro científico en que *esa* clase de libre albedrío es una ilusión, pero eso no significa que el libre albedrío sea una ilusión en ningún sentido relevante desde el punto de vista moral. Es tan real como los colores, tan real como los dólares.

Desgraciadamente, algunos de los científicos que ahora declaran que la ciencia ha demostrado que el libre albedrío es una ilusión pasan a decir que este «descubrimiento» es importante en un sentido relevante desde el punto de vista moral. Piensan que tiene enormes implicaciones para la moral y la ley: nadie es *realmente* responsable, por ejemplo, así que nunca nadie *merece* ser ni castigado ni elogiado. Están cometiendo el mismo error de la gente que dice que nada es nunca sólido, no *realmente*. Están empleando un concepto popular recalcitrante del libre albedrío, cuando primero deberían ajustarlo, tal como hacen con el color y la conciencia (y el espacio, el tiempo, la solidez y todas las otras cosas en las que la *ideología* de la imagen manifiesta se equivoca).

Las bombas de intuición de esta parte están concebidas para alejarte de esa ideología sobre el libre albedrío y persuadirte de ver un mejor concepto, el concepto del libre albedrío real, el libre albedrío práctico, el fenómeno de la imagen manifiesta que importa. Demasiadas polémicas han girado alrededor del tema del libre albedrío a lo largo de varios siglos, y están demasiado enmarañadas para resolverlas en un capítulo de un libro o en el libro entero, pero en algún lugar tenemos que empezar, y estas herramientas de pensamiento funcionan un poco como palancas que te sacan de surcos ya muy gastados y te llevan a un terreno bastante nuevo con mejores perspectivas. La primera está pensada para mostrar por qué ésta es una tarea tan importante.

65. Una neurocirujana verdaderamente nefanda

Estamos en los albores del tratamiento neuroquirúrgico de afecciones psicológicas debilitantes. La estimulación profunda del cerebro con implantes de electrodos está mostrando efectos asombrosos en el tratamiento del trastorno obsesivo compulsivo, por ejemplo, según se informa en un estudio pionero que llevaron a cabo el neuropsiquiatra Damiaan Denys y sus colegas en Ámsterdam.[1] Eso es un hecho, pero esto es ficción: un buen día, una brillante neurocirujana le dijo a un paciente al que acababa de realizarle un implante en su flamante teatro de operaciones de alta tecnología:

> El dispositivo que te implanté no sólo controla tu trastorno obsesivo compulsivo: controla todas tus decisiones gracias a nuestro sistema de control maestro, que mantiene contacto por radio con tu microchip las veinticuatro horas. En otras palabras, he desarticulado tu libertad consciente; a partir de ahora tu sensación de libre albedrío será una ilusión.

Lo cierto es que no había hecho nada de eso; era simplemente una mentira que había decidido decirle a su paciente para ver qué pasaba. Funcionó: el pobre tipo salió al mundo convencido de que no era un agente responsable, sino más bien un simple títere, y se empezó a notar en su comportamiento: se volvió irresponsable, agresivo y negligente, y se consentía sus peores caprichos, hasta que lo atraparon y llevaron a juicio. Testificando en su propia defensa, ardientemente argumentaba que debido al implante en su cerebro no era reponsable, y la neurocientífica, cuando la llamaron a testificar, reconoció lo que había dicho, y añadió: «Pero sólo estaba tomándole el pelo; no fue más que una broma. ¡Nunca pensé que me creería!»

Realmente no importa si el tribunal creyó el testimonio de él o el de ella, si lo sentenció a él o a ella. Sea como sea, ella arruinó la vida de él con su poco meditada afirmación: lo privó de su integridad y mutiló su poder de tomar decisiones. De hecho, el falso «informe» sobre la operación que le dio a su paciente consiguió de un modo no quirúrgico mucho más de lo que afirmaba haber conseguido con la cirugía: lo incapacitó. Pero si ella es responsable de esta grave consecuencia, los neurocientíficos que hoy en día ocupan los medios de comunicación para hablar de cómo su ciencia muestra que el libre albedrío es una ilusión están arriesgándose a provocar en masa el mismo daño a toda la gente que les cree a pie juntillas.[2] Los neurocientíficos, los psicólogos y los filósofos necesitan tomarse en serio su obligación moral de considerar detenidamente las presuposiciones e implicaciones de sus declaraciones públicas sobre estos temas, y con el mismo cuidado que se le exige a la gente que pontifica sobre el calentamiento global o los inminentes impactos de asteroides. Para dar un solo ejemplo, considera el mensaje que el

astuto crítico social y observador Tom Wolfe encuentra en las declaraciones de estos neurocientíficos:

> La conclusión que está sacando la gente más allá de los muros del laboratorio es: ¡Es *un complot!* ¡*Todos estamos programados!* Eso y: ¡*No me eches la culpa!: estoy mal cableado.*[3]

¿Mal cableado? ¿Entonces qué sería estar bien cableado, o que los científicos «descubrieran» que nadie está, ni podría estar, bien cableado para la responsabilidad moral?

66. Un juguete determinista: el juego de la vida de Conway

El físico Richard Feynman tenía una pregunta favorita para plantearles a los oradores cuando se encontraba en una conversación científica sobre un tema que no conocía tan bien: ¿me puedes dar un ejemplo sencillísimo de lo que estás diciendo? Si el orador no podía complacerlo, Feynman desconfiaba, y con toda razón. ¿De verdad esta persona tenía algo que decir, o no eran más que florituras técnicas haciéndose pasar por sabiduría científica? Si no puedes explicar de manera relativamente simple un problema difícil, probablemente no lo estás haciendo bien. La simplificación no es sólo para principiantes.

La biología tiene sus organismos modelo, especies que han sido meticulosamente elegidas para facilitarle el camino al experimentador; se reproducen rápido en el laboratorio, son relativamente seguras y fáciles de manejar, y —una vez que muchos equipos las han estudiado— se ha hecho su cartografía genética y se ha llegado a entenderlas muy bien: la mosca de la fruta, la rata de laboratorio, el pez cebra, el calamar (por su axón gigante), el nematodo *Caenorhabditis elegans* y la *Arabidopsis thaliana,* una planta emparentada con la mostaza, que es resistente, crece muy rápido y fue la primera planta con el genoma completo secuenciado. La inteligencia artificial tiene sus propios casos sencillos, conocidos como «problemas de juguete», que, como el nombre sugiere, son versiones de problemas «serios» del mundo real que de forma deliberada se han simplificado exageradamente. Muchos de los programas más interesantes que se han ideado para la inteligencia artificial son soluciones a problemas de juguete, como hacer que un programa de computadora construya estructuras simples del *mundo de los bloques,* un mundo virtual que consiste en un tablero con un montón de cubos para niños. Jugar al ajedrez es un problema de juguete; seguramente es mucho más manejable que conducir de Maine a California, que resolver el conflicto árabe-israelí o incluso que encontrar en la cocina ingredientes apropiados para un sándwich y prepararlo. Los expertos en ética tienen sus *dilemas del tranvía.* En su versión más simple, un tranvía fuera de control está descendiendo por una vía donde, si no cambia de dirección, golpeará y matará a cinco personas que no pueden quitarse de su paso. Hay un cambio de agujas que puedes mover para desviar el tranvía a una vía lateral…, donde golpeará y matará a una sola persona. ¿Lo mueves?

Hay también un mundo de juguete para ayudar a la gente a pensar sobre el *determinismo,* la idea de que los hechos en un momento dado en el tiempo —la ubicación, la masa, la dirección, la velocidad de todas las partículas— determina lo que pasa en el momento siguiente, y así sucesivamente, para siempre. Durante miles de años, los físicos y los filósofos, entre otros, han discutido si nuestro universo es determinista o si hay algunos sucesos genuinamente no determinados: sucesos «azarosos» del todo impredecibles que simplemente *ocurren* sin que nada los cause. Hasta los pensadores

experimentados pueden encontrar muy estimulante juguetear con Life [o el juego de la vida —por *life,* «vida» en inglés—], el modelo de un mundo determinista, de una sencillez pasmosa, creado por el matemático John Horton Conway y sus alumnos de posgrado en 1970.

Life se juega en una cuadrícula de dos dimensiones, como un tablero de damas, con fichas simples, como piedritas o monedas —alguien podría también verse muy tecnológico y jugarlo en una pantalla de computadora—. No es un juego en el que se gane o pierda; si es que es un juego, es solitario. Cada una de las casillas cuadradas en que se divide el plano de la cuadrícula está a cada momento prendida (ON) o apagada (OFF) (si está ON pones una ficha en el cuadro; si está OFF, dejas el cuadro vacío). Observa que cada casilla tiene ocho vecinos: las cuatro casillas adyacentes —norte, sur, este y oeste— y las cuatro diagonales —noreste, sureste, suroeste y noroeste—.

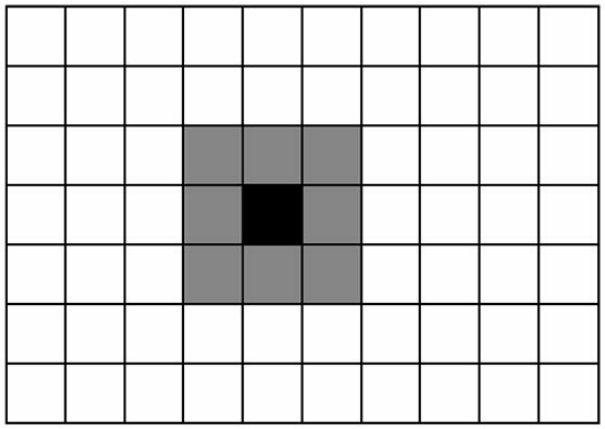


FIGURA 17

En Life el tiempo es discontinuo, no continuo; avanza en momentitos, y el estado del mundo cambia entre un momentito y otro de acuerdo con la siguiente regla:

> *Física de Life:* para cada casilla de la cuadrícula, cuenta cuántos de sus ocho vecinos están ON en el instante presente. Si la respuesta es exactamente dos, en el siguiente instante la casilla se quedará en su estado presente (ON u OFF). Si la respuesta es exactamente tres, la casilla estará ON en el instante siguiente, sea cual sea su estado actual. En todas las demás condiciones, la casilla estará OFF.

Ésa es la única regla del juego. Ya sabes todo lo que necesitas para poder jugar Life. *Toda la física del mundo de Life se recoge en esa sencilla ley sin excepciones.* Aunque ésta es la ley fundamental de la «física» del mundo de Life, al principio es útil a concebir esta física curiosa en términos biológicos: piensa en las casillas que se ponen en ON como nacimientos, en las casillas que se ponen en OFF como muertes, y los instantes subsiguientes como generaciones. Tanto el hacinamiento (tener más de tres casillas

vecinas habitadas) como el aislamiento (menos de dos casillas vecinas habitadas) llevan a la muerte. Analiza unos cuantos casos.

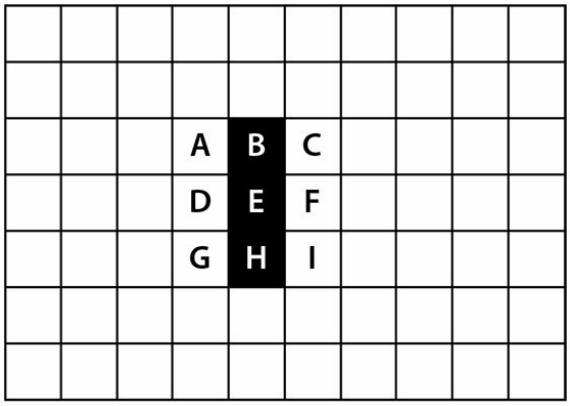


FIGURA 18

En la configuración de la figura 18, sólo las casillas D y F tienen exactamente tres vecinos ON, así que serán las únicas casillas de nacimiento en la siguiente generación. Las casillas B y H tienen sólo un vecino ON, así que morirán en la siguiente generación. La casilla E tiene dos vecinos ON así que se queda encendida. Entonces, el siguiente «instante» se verá como la figura 19.

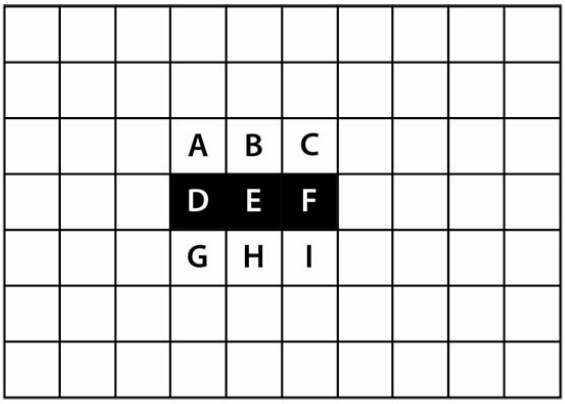


FIGURA 19

Evidentemente, la configuración se revertirá en el siguiente instante, y este pequeño motivo cambiará de ida y vuelta indefinidamente, a menos que de algún modo entren en escena algunas nuevas casillas ON. Se lo llama un *intermitente* o semáforo. ¿Qué le pasará a la configuración de la figura 20?

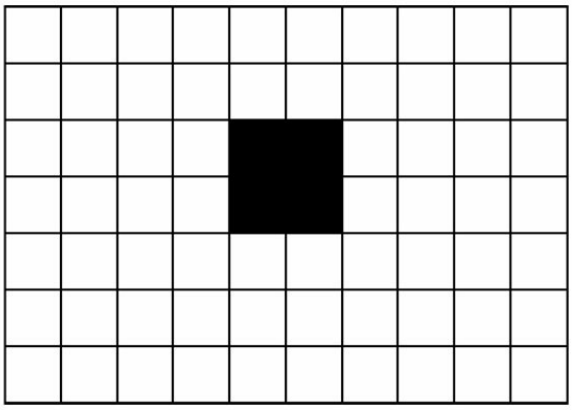


FIGURA 20

Nada. Cada casilla ON tiene tres vecinos ON, así que renace tal como es. Ninguna casilla OFF tiene tres vecinos ON, así que no habrá más nacimientos. A esta configuración se la llama *naturaleza muerta.*

Mediante la aplicación escrupulosa de nuestra única ley, alguien puede predecir con total exactitud el siguiente instante de cualquier configuración de las casillas ON y OFF, y el instante siguiente, y así sucesivamente. *En otras palabras, el mundo de Life es un mundo de juguete que ejemplifica a la perfección el determinismo que Pierre Laplace, científico francés de principios del siglo XIX, hizo famoso: dada la descripción del estado de este mundo en este instante, los observadores podemos predecir perfectamente los instantes futuros mediante la simple aplicación de nuestra* única *ley de la física.* O podríamos decirlo de este otro modo: cuando *adoptamos la postura física* frente a una configuración del mundo de Life, nuestros poderes predictivos son perfectos: no hay ruido, no hay incertidumbre, no hay probabilidad de menos de uno. Es más, del hecho de que el mundo de Life tenga dos dimensiones se sigue que no hay nada escondido. No hay nada tras bambalinas, no hay variables ocultas; todo lo que ocurre con la física de los objetos del mundo de Life se puede ver directamente.

Si seguir la simple regla te resulta un ejercicio tedioso, hay simulaciones por computadora del mundo de Life en las que puedes crear configuraciones en la pantalla y dejar que la computadora ejecute el algoritmo por ti, y cambie la configuración una y otra vez según la regla. En las mejores simulaciones puede cambiarse la escala tanto de tiempo como de espacio, y verlo alternadamente en primer plano y a vuelo de pájaro.

Pronto se descubre que algunas configuraciones sencillas ON son más interesantes que otras. Por ejemplo, un segmento en diagonal, como el que se muestra en la figura 21.

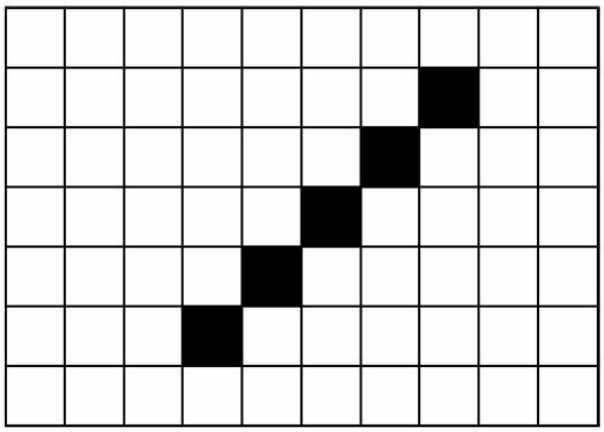


FIGURA 21

Este motivo *no* es un intermitente; cada generación, las dos casillas en ON a los extremos de la fila mueren de aislamiento y no hay casillas de nacimiento. Pronto se evapora todo el segmento. Aparte de las configuraciones que nunca cambian —las naturalezas muertas— y las que se evaporan por completo, como el segmento en diagonal, hay otras con todo tipo de periodicidad. Como vimos, el intermitente tiene un periodo de dos generaciones que continúa *ad infinitum,* a menos que otra configuración llegue a invadir la cuadrícula. Las invasiones son lo que hace a Life interesante: entre las configuraciones periódicas hay algunas que nadan como amibas a través del plano. La más simple es el *planeador;* la configuración de cinco pixeles que aparece moviéndose hacia el sureste con una sola brazada en la figura 22.

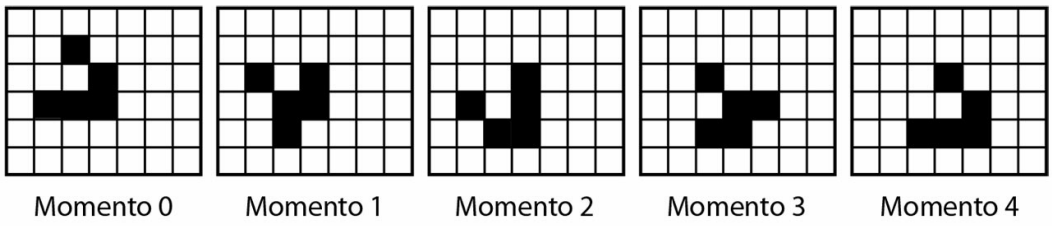


FIGURA 22

Y luego tenemos a los tragones, los trenes de vapor, los rastrillos espaciales y una gran cantidad de habitantes del mundo de Life con nombres apropiados que surgen como objetos reconocibles en un nuevo nivel (análogos al *nivel de diseño*). Este nivel tiene su propio lenguaje, un escorzo transparente de las tediosas descripciones que podrían darse en el nivel físico. Por ejemplo:

> Un tragón puede comer un planeador en cuatro generaciones. Se consuma lo que se consuma, el proceso básico es el mismo. Se forma un puente entre el tragón y su presa. En la siguiente generación, la zona del puente muere de sobrepoblación y da

una mordida tanto al tragón como a la presa. Entonces el tragón se repara; la presa normalmente no puede. Si lo que queda de la presa muere como pasó con el planeador, la presa se consume.[4]

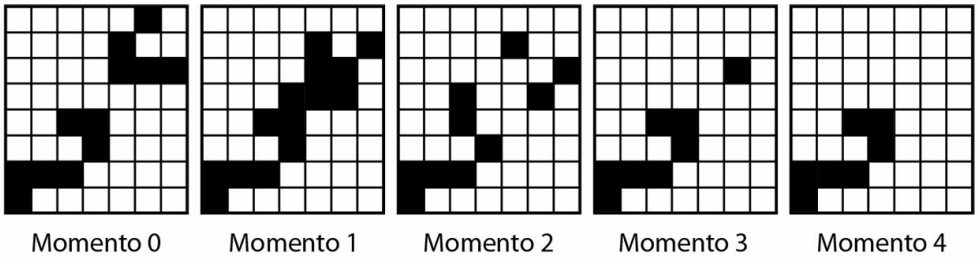


FIGURA 23

Observa que algo curioso le pasa a nuestra «ontología» —nuestro catálogo de lo que existe— mientras nos movemos entre niveles. En el nivel físico no hay movimiento, sólo ON y OFF, y las únicas cosas individuales que existen, las casillas, se definen por su posición espacial fija. En el nivel de diseño de pronto tenemos el movimiento de los objetos persistentes; este planeador y este otro son el mismo (aunque cada generación se compone de diferentes casillas) que en la figura 22 se ha movido hacia el sureste, cambiando de forma mientras se mueve; y en la figura 23 hay un planeador menos en el mundo después de que el tragón se lo ha comido.

Observa también que mientras que en el nivel físico no hay absolutamente ninguna excepción a la ley general, en este nivel nuestras generalizaciones requieren cláusulas que las acoten: «por lo general», «siempre y cuando nada invada»… Algunos restos perdidos de sucesos anteriores pueden «romperse» o «matar» a algún objeto de la ontología en este nivel. Su *prominencia como cosas reales* es considerable, pero no está garantizada. Decir que su prominencia es considerable es decir que, corriendo un pequeño riesgo, puede ascenderse a este nivel de diseño, adoptar su ontología y pasar a predecir —de manera superficial y un poco arriesgada— el comportamiento de configuraciones o sistemas de configuraciones más grandes, sin tomarse la molestia de computar el nivel físico. Por ejemplo, podemos asignarnos la tarea de diseñar un interesante supersistema a partir de las «partes» que el nivel de diseño pone a nuestra disposición.

Esto es exactamente lo que Conway y sus alumnos se propusieron, y lo consiguieron majestuosamente. Diseñaron una entidad autorreproductora compuesta tan sólo de casillas de Life, y demostraron la viabilidad de su diseño. Dándole duro de manera determinista en su plano infinito, se copiaba a sí misma a la perfección, y luego esta copia se copiaba a sí misma, y así sucesivamente. Por si fuera poco, esto era además una máquina de Turing universal: ¡una computadora de dos dimensiones que en principio puede computar cualquier función computable! ¿Qué demonios movió a Conway y a sus

alumnos a crear primero ese mundo y luego este maravilloso habitante de ese mundo? Estaban tratando de responder, en un nivel muy abstracto, una de las principales preguntas de la biología: ¿cuál es la complejidad mínima necesaria para que una cosa se pueda autorreproducir? Estaban continuando las brillantes especulaciones de John von Neumann, quien cuando murió en 1957 estaba trabajando sobre esa pregunta. Francis Crick y James Watson habían descubierto el ADN en 1953, pero cómo funcionaba siguió siendo un misterio muchos años. Von Neumann había imaginado con cierto detalle una especie de robot flotante que recogía desechos que podían usarse para armar un duplicado de sí mismo, y que además pudiera repetir el proceso. Su descripción (publicada póstumamente en 1966) de cómo un autómata leería su propio plano y luego lo copiaría en su nueva creación previó con impresionante detalle muchos de los futuros descubrimientos sobre los mecanismos de expresión y copia del ADN, pero a fin de que su prueba de la posibilidad de un autómata que se reproduce a sí mismo fuera matemáticamente rigurosa y manejable, Von Neumann cambió a abstracciones sencillas, de dos dimensiones, ahora conocidas como *autómatas celulares*. Las casillas del mundo de Life de Conway son un ejemplo especialmente simpático de autómata celular.

Conway y sus alumnos querían confirmar en detalle la prueba de Von Neumann mediante la creación de un verdadero mundo de dos dimensiones con una física simple en la cual la construcción autorreproducible fuera una estructura estable y *funcional*. Como Von Neumann, querían que su respuesta fuera lo más general posible, y en consecuencia tan independiente como se pudiera de la física y la química reales (¿terrenales?, ¿locales?). Querían algo de lo más simple, fácil de visualizar y fácil de calcular, así que no sólo bajaron de tres dimensiones a dos, sino que también «digitalizaron» tanto el espacio como el tiempo: todos los tiempos y distancias, como vimos, están en números enteros de «instantes» y «casillas». Era Von Neumann quien había tomado la concepción abstracta de una computadora mecánica de Alan Turing (ahora llamada máquina de Turing); con ella elaboró las especificaciones de una computadora para todo uso con programas almacenados y procesamientos en serie (ahora llamada máquina de Von Neumann), y en sus brillantes exploraciones de los requisitos espaciales y estructurales para una computadora así, se dio cuenta de que una máquina de Turing universal (véase el capítulo IV) podría en principio «construirse» en un mundo de dos dimensiones —y lo demostró—.[5] Conway y sus alumnos también se propusieron confirmarlo con su propio ejercicio de ingeniería en dos dimensiones.[6]

No fue fácil ni mucho menos, pero mostraron cómo podían «construir» una computadora funcional a partir de formas de Life más simples. Flujos de planeadores podrían, por ejemplo, proporcionar la «cinta», y el lector de cintas podía ser una enorme concurrencia de tragones, planeadores y otras cosas. ¿Qué aspecto tiene esta máquina? Poundstone calculó que toda la construcción tendría alrededor de 10^{13} casillas o pixeles.

> Para visualizar en la pantalla un dibujo de 10^{13} pixeles se necesitaría una pantalla de video de al menos unos 3 millones de pixeles a lo ancho […] [Imagina una pantalla con la alta resolución de tu computadora portátil o tu iPad, de unos 800 metros de ancho.] La perspectiva reduciría a la invisibilidad los pixeles de un dibujo que se autorreprodujera. Si te alejas de la pantalla lo suficiente para que todo el patrón esté cómodamente a la vista, los pixeles (e incluso los planeadores, tragones y pistolas) serían demasiado pequeños para distinguirse. Un dibujo autorreproductor sería un resplandor brumoso, como una galaxia.[7]

En otras palabras, para cuando hayas acumulado suficientes piezas para construir algo que pueda reproducirse a sí mismo (en un mundo de dos dimensiones), será, a grandes rasgos, tanto más grande que sus pedazos más pequeños como un organismo lo es con respecto a sus átomos. Probablemente no puedas hacerlo con nada mucho menos complicado, aunque esto no se ha demostrado totalmente.

Life, o el juego de la vida, ilustra muchos principios importantes y puede usarse para elaborar diferentes argumentos o experimentos mentales, pero yo lo usaré para ilustrar sólo tres cuestiones, y dejo que tú descubras otras por ti solo. En primer lugar, observa cómo aquí se difumina la distinción entre la postura física y la postura de diseño. Por ejemplo, ¿los planeadores cuentan como cosas diseñadas o como objetos naturales (como los átomos y las moléculas)? Si algo puede contar como cosa diseñada es el lector de cintas que Conway y sus alumnos hicieron juntando planeadores, tragones y cosas por el estilo, pero sus ingredientes son materias primas muy crudas: las «cosas» más simples en el mundo de Life. Nadie tuvo que diseñar o inventar al planeador: se *descubrió* que la física del mundo de Life lo implicaba. Eso, sin embargo, realmente puede decirse de *todo* lo que hay en el mundo de Life. Nada pasa en ese mundo que no se implique rigurosamente —que no sea deducible lógicamente por sencillas demostraciones de teoremas— de su física y de las configuraciones iniciales de las casillas. Simplemente, algunas de las cosas del mundo de Life son más maravillosas e imprevistas (al menos nosotros, con nuestros raquíticos intelectos, no podemos preverlas) que otras. Hay un sentido en el que la galaxia de pixeles de computadora que se autorreproducen concebida por Conway es «sólo» otra macromolécula de Life con una periodicidad de comportamiento muy larga y complicada. Esto ilustra muy bien un tema paralelo de biología y el origen de la vida: podría decirse que los aminoácidos simplemente son; no tuvieron que diseñarse. Sin embargo, las proteínas, compuestas exclusivamente de aminoácidos, son demasiado elaboradas; son, al menos, algo **maomeno** diseñado. El gradualismo de Darwin hace una nueva aparición.

En segundo lugar, el mundo de Life, al ser determinista, tiene un futuro perfectamente predecible para cualquier posible configuración, pero, de manera un poco

sorprendente, su pasado es a menudo perfectamente inescrutable. Piensa en la naturaleza muerta que consiste en cuatro pixeles ON formando un cuadrado. No puedes saber, sólo por verla, o incluso por verlas a ella y a su vecindario, cuál fue su pasado. Para entender esto, observa que cualesquiera tres de los cuatro pixeles ON conducirían, en la siguiente generación, a esta naturaleza muerta de cuatro pixeles ON. Si alguna de esas células estuvo OFF en el pasado es un **hecho histórico inerte**.

En tercer lugar, recuerda lo importantes que eran el «ruido» y las colisiones para crear las mutaciones de las que la evolución —al igual que otros procesos creativos— se alimenta. La inmensa construcción de Conway se reproducía a sí misma, pero no podía mutar. Siempre haría una copia *perfecta* de sí misma, y para que las mutaciones puedan entrar en escena, toda la construcción tendría que agrandarse muchas veces. ¿Por qué? Porque el mundo de Life es determinista, así que la única manera como puede ocurrir una mutación aleatoria es si algún pedazo perdido y errabundo entra (seudoaleatoriamente) en escena y rompe algo. Sin embargo, la cosa moviente más pequeña es un planeador, así que imagínalo como un solo fotón, o un rayo cósmico, moviéndose a la velocidad de la luz (de la física de Life). Un solo planeador puede hacer un gran daño; si no tiene más remedio que apenas «retocar» algo en el genoma de la cosa autorreproductiva sin destruir el genoma, ese genoma tendrá que ser muy grande con relación al planeador, y de lo más resistente. Sería posible demostrar que la evolución no podría ocurrir en el mundo de Life, por grandes que hiciéramos las entidades, si resultara que estos ensamblajes del tamaño de una galaxia son demasiado frágiles para sobrevivir a una lluvia ocasional de planeadores.

67. Piedra, papel o tijeras

Probablemente todos mis lectores conozcan el juego de «piedra, papel o tijeras». Dos personas, una enfrente de la otra, dicen «Piedra, papel o tijeras», y simultáneamente muestran la mano, ya sea cerrada en un puño (piedra), con dos dedos extendidos (tijeras) o con la palma abierta hacia abajo (papel). La piedra rompe las tijeras (y les gana), las tijeras cortan el papel (y le ganan) y el papel envuelve la piedra (y le gana). A menos que los dos jugadores muestren el mismo símbolo, un jugador gana y otro pierde. Es un juego seductor porque si puedes adivinar lo que hará tu oponente, puede parecer que lees su mente y tú consistentemente sales con la forma que le va a ganar. Es desconcertante cuando eso pasa. ¿Algunas personas pueden jugar mejor que otras? Por lo visto sí, pues ha habido torneos con grandes premios en metálico de los que han surgido ganadores nacionales e internacionales y —esto es importante porque en todo terreno alguien tiene que ganar aunque no haya habilidades implicadas— los mejores jugadores tienen un historial de victorias.

¿Cómo lo hacen? A lo mejor detectan sutiles señales en las caras y posturas de sus oponentes. Los jugadores de póquer dicen que interpretan lo que otro jugador «delata», sienten cuando están «blofeando» o faroleando y cuando no, sin dejar de poner su propia «cara de póquer». A lo mejor la mayoría de la gente que juega piedra, papel o tijeras tiene algo que lo delata y que no puede controlar, y que en el último instante detectan los jugadores. ¿Entonces cuál es la mejor estrategia para impedir que tu oponente recoja una regularidad en tus expresiones externas? Jugar de manera completamente aleatoria es lo mejor, pues si juegas una secuencia aleatoria de movimientos, simplemente no habrá una regularidad que tu oponente pueda detectar. (Mientras juegas aleatoriamente, y los dos contrincantes van ganando parejo, puedes tratar de encontrar una regularidad en los movimientos del otro, y usarla para trabajar una estrategia no aleatoria, y luego dar el golpe.)

La gente es notoriamente mala cuando se pone a crear series en efecto aleatorias. Tratan de cambiar todo el tiempo y evitan elegir el mismo movimiento dos o tres veces seguidas, por ejemplo (algo que debería ocurrir bastante a menudo en una serie genuinamente aleatoria). Al saber que tu esfuerzo improvisado por crear una serie genuinamente libre de regularidades puede fallar, debes considerar una mejor estrategia: consigue en la biblioteca (o en línea) una tabla de números aleatorios. Pon el dedo «aleatoriamente» en algún punto de la tabla, y copia los siguientes 100 dígitos. Deshazte de todos los ceros (digamos) y remplaza cada 1, 2 y 3 con una «R» (para «piedra»); cada 4, 5 y 6 con una «P» (de «papel») y cada 7, 8 y 9 con «T» (de «tijeras»). Esto te dará una secuencia de aproximadamente 90 movimientos (pues te deshiciste de aproximadamente 10 ceros), que debería ser suficiente para un torneo.

Ahora estás listo para jugar, y la regla fundamental es: *mantén en secreto tu lista.* Si

tu oponente le da un vistazo, estarás completamente a su merced. Te convertirá en una fábrica de dinero, como se dice. Si no hay modo de que tenga acceso a tu lista, por otra parte, tendrá que empezar a tratar de anticipar tu reacción, previendo tu línea de pensamiento lo mejor que pueda. (En pocas palabras, tendrá que tratarte desde la postura intencional y razonar sobre tu razonamiento, en lugar de tratarte como un simple mecanismo cuyo comportamiento se puede conocer en el manual…, o en la lista).

Este sencillo principio de ocultarle a tu oponente las jugadas que harás resulta ser uno de los principales ejes alrededor de los cuales giran desde hace tiempo las polémicas sobre el libre albedrío. De hecho, la invención de la *teoría de juegos* por Von Neumann y Morgenstern comenzó con el reconocimiento de que mientras un agente solitario (o sistema intencional), tratando de predecir el futuro de la información que recoge, puede arreglárselas con un cálculo de *utilidad esperada* usando la teoría de la probabilidad (de una u otra clase), en cuanto hay en el medio ambiente dos agentes, dos sistemas intencionales, la circunstancia cambia radicalmente.[8] Ahora cada agente tiene que tomar en cuenta los intentos de predicción del otro agente, e incluir la observación de su propio comportamiento que hace ese otro agente, así como sus intentos de preverla y aprovecharla, con lo que se crean bucles de realimentación de una complejidad indefinida.[9]

Esta niebla fundamental de inescrutabilidad o impredecibilidad entre los agentes crea las condiciones en las que florece la teoría de juegos. La evolución lo ha descubierto, y se puede ver a muchas especies aplicando los principios de la teoría (**¡competencia sin comprensión!**) a sus interacciones. Los **brincos de las gacelas** son un ejemplo sencillo. Otro es el vuelo errático de las mariposas, cuyas trayectorias son difíciles de predecir para los pájaros insectívoros, aunque la evolución ha ayudado a los pájaros al darles un «umbral de fusión de parpadeo» mayor que el nuestro, por ejemplo (ellos ven más «cuadros por segundo» que nosotros, y una película sería para ellos como una proyección de diapositivas).

Sé impredecible y busca a otros que sigan este mismo consejo. La «valoración» de este principio puede verse en el generalizado instinto que se encuentra en animales que, cuando se enfrentan a una entidad de movimientos complicados, intentan tratarla como agente —«¿*Quién* anda ahí y qué quiere?», no sólo «¿Qué fue eso?»— a fin de estar a salvo, porque a lo mejor *sí* es un agente y quiere comerse al animal, o aparearse con él, o disputarse algún premio. Esta respuesta instintiva es, en la evolución, lo que explica que se hayan inventado todos los duendes, elfos, aluxes, hadas, ogros y dioses invisibles que a la larga evolucionaron hasta convertirse en Dios, el supremo sistema intencional invisible.[10]

Como los animales que adoptan esta estrategia sin tener que saber por qué, los seres humanos valoramos la táctica de conservar nuestra impredecibilidad sin tener que

entender por qué es tan buena idea. Sin embargo, la mayor parte del tiempo es de lo más evidente. Cuando vas al tianguis y detectas una antigüedad que debes tener a como dé lugar, sabes bien que no te conviene que se te note la emoción antes de que te hayan dicho cuánto cuesta. Si no lo haces así, el vendedor le inflará el precio. Cuando anuncias algo que has puesto a la venta, pones un precio de salida razonable —uno que estarías satisfecho de recibir— porque no puedes saber cuándo alguien estaría dispuesto a pagar más, y esperas que el comprador no sepa cuánto menos que tu precio de salida aceptarías (y desde luego que puedes negarte a regatear). Las subastas son formas de explorar esta *terra incognita,* y si le confías al subastador tu puja, cuentas con su integridad para que no les revele tu máximo a los otros postores.

Del mismo modo, cuando te enamoras locamente a primera vista de alguien, haces tu mejor esfuerzo para no desvanecerte y jadear, y por mantenerte tan reservado e indiferente como sea posible. No quieres espantar a este dechado de virtudes, pero tampoco quieres que te domine y te eche a la bolsa. Una cara de póquer no es sólo para el póquer. Por lo general, aumentas tus probabilidades de conseguir lo que quieres si mantienes desconcertada a la competencia —los otros agentes que hay en los alrededores —, para que no tengan suficientes corazonadas sobre dónde vas a saltar y entonces no puedan prepararse para tu aterrizaje (es costoso preparar el entorno de otro agente, así que tus oponentes no tratarán de predecirte a menos que tengan buenos indicios de lo que vas a hacer).

Los magos saben cómo ejercer una «fuerza psicológica» para hacer que en una baraja escojas («haciendo uso de tu libre albedrío») la carta que quieren que tomes. Hay muchos métodos, todos ellos sutiles y difíciles de detectar, y un buen mago puede tener éxito casi siempre. Esto es un genuino encogimiento de tu agencia, una manipulación que te convierte en una herramienta, un títere, una extensión de la voluntad del mago, no en un agente libre.

A diferencia de lo que dice la ideología antigua, no queremos que nuestras elecciones libres carezcan totalmente de una causa. Lo que todos queremos, y debemos querer, es que cuando actuamos lo hagamos basados en buena información sobre las mejores opciones disponibles. ¡Ay, si tan sólo el entorno *causara* que tuviéramos un montón de creencias verdaderas pertinentes sobre lo que está allá afuera, y también *causara* que actuáramos a partir de la más acertada evaluación de esas señales que pudiéramos obtener! Eso nos daría *casi* todo lo que como agentes queremos, excepto esto: no querríamos que en el entorno hubiera un agente manipulador que nos usurpara el control, así que no querríamos que el entorno hiciera nuestras mejores jugadas muy evidentes a todos los otros agentes que están allá afuera, pues entonces, al saber demasiado sobre lo que queremos y con cuántas ganas lo queremos, podrían aprovecharse de nosotros. Así que añade a nuestra lista de deseos la capacidad de mantener en secreto nuestros

procesos mentales y nuestras decisiones, aunque eso en ocasiones signifique que elegimos no nuestra mejor opción sino la que queda en segundo lugar, sólo para mantener a los demás en desconcierto.[11]

Algunas personas, que vagamente aprecian la importancia de esta impredecibilidad, piensan que «para ir a lo seguro» deberían tener una impredecibilidad *absoluta,* que sólo puede lograrse si en el sótano de nuestros cerebros los asuntos son físicamente indeterministas. Así lo dijo una vez el filósofo Jerry Fodor con su característica viveza: «Uno quiere ser lo que dice la tradición que era Eva cuando mordió la manzana: perfectamente libre de no hacerlo. Tan perfectamente libre, de hecho, que ni siquiera Dios sabía de qué lado se inclinaría».[12] Pero ¿por qué «uno quiere» esto? ¿De verdad esta absoluta impredecibilidad es preferible a una impredecibilidad práctica? Muchos filósofos, a lo largo de varios miles de años, han insistido en la impredecibilidad absoluta como condición de un genuino libre albedrío. ¿Saben algo que los demás no sabemos? Si es así, lo están manteniendo en secreto. A casi todos ellos, la idea de que el libre albedrío es incompatible con el determinismo les ha parecido demasiado evidente para necesitar un argumento. Yo digo que ellos tienen la carga de la prueba. Que nos muestren por qué deberíamos estar desesperados si no podemos tener impredecibilidad *absoluta.* He mostrado por qué para nosotros —como para la evolución misma— es prudente tratar de conseguir toda la impredecibilidad *práctica* que se pueda. Que nos digan, por favor, por qué no es suficiente.

Te voy a dar una buena razón: si planeas jugar piedra, papel o tijeras con Dios y hacer apuestas muy altas (la salvación, tal vez), entonces tienes razón para querer una libertad «perfecta», como dice Fodor. Al menos yo no pienso que esa competencia vaya a tener lugar, así que estoy satisfecho con la libertad práctica que tengo con tan sólo seguir mi consejo y, cuando hay mucho en juego, mantenerme lejos de los artistas de la prestidigitación y otros manipuladores.

68. Dos loterías

Compara la imparcialidad de las siguientes dos loterías. En la lotería A ya se vendieron todos los billetes, sus talones están en una mezcladora y se mezclan, tan aleatoriamente como quieras, y *luego* se saca al azar el billete ganador (la mayoría de las loterías conocidas son así). En la lotería B, la mezcla de los talones y la extracción al azar del ganador tienen lugar *antes* de que se vendan los billetes (y el talón ganador se pone en una caja de seguridad), pero en todo lo demás ambas loterías se realizan igual. Alguien podría pensar que la segunda lotería es parcial porque el billete ganador se determinó antes de que la gente siquiera comprara sus billetes. Uno de esos billetes *ya* es el ganador (aunque nadie sepa cuál es); los otros billetes son papel sin ningún valor, y vendérselos a la gente desprevenida es una especie de fraude. Sin embargo, de hecho las dos loterías son igualmente imparciales. Todos los que compran un billete tienen las mismas probabilidades de ganar. El momento en que se seleccione al ganador no es en lo absoluto una característica esencial.

En la mayoría de las loterías se pospone el momento de sacar el talón hasta que se han vendido todos los billetes, a fin de proporcionarle al público la posibilidad de ser testigos de primera mano de que no ha habido chanchullos. Ninguna persona mañosa con conocimiento confidencial ha manipulado la distribución de los billetes, porque el conocimiento del billete ganador no existe (ni podría existir) en ningún agente hasta que los billetes se han vendido. Es interesante que no todas las loterías sigan esta práctica. La compañía Publisher's Clearing House solía mandar cada año por correo millones de sobres que decían con letras grandes «puede ser que ya seas el ganador» —de un millón de dólares o de algún otro premio—. (Sigue organizando una lotería, pero sobre todo por internet.) Estas costosas campañas están basadas en investigación de mercado que muestra que la gente por lo general sí piensa que las loterías con ganadores preseleccionados son imparciales y justas siempre y cuando se realicen con honestidad, aunque tal vez las acepta sin quejarse porque los billetes le salen gratis. ¿Mucha gente *compraría* un billete en una lotería en la que el talón ganador estuviera en un sobre lacrado y se supiera que desde el principio se depositó en una bóveda de seguridad? La gente compra millones de boletos de raspa y gana, y en el momento de la compra ya ha sido determinado si algún boleto es ganador. Por lo visto esta gente piensa que tiene una oportunidad real de ganar. Pienso que tiene razón, pero sin importar si la tiene o no, su tranquila convicción de que esas loterías son imparciales y de que tiene una oportunidad real de ganar debería anular la confianza de los filósofos (que se remonta 2 000 años atrás, a Epicuro y Lucrecio) que de algún modo se han convencido a sí mismos de que ninguna oportunidad es una oportunidad *real* a menos que el resultado se determine sólo en el último instante. Estos filósofos han sostenido que sin un continuo suministro de *puntos de ramificación* indeterminados que rompan la tela de la causalidad no hay

posibilidad de que haya elecciones verdaderas, no hay oportunidades *reales* de hacer las cosas bien.

Las dos loterías nos permiten mirar el determinismo desde una nueva perspectiva. Si el mundo está determinado, entonces tenemos adentro de nosotros generadores de números seudoaleatorios, no aleatorizadores verdaderamente aleatorios (cuánticos). Si nuestro mundo está determinado, todos nuestros billetes de lotería se extrajeron, en efecto, al mismo tiempo, hace 14 000 millones de años en el momento del Big Bang, se metieron en un sobre a nuestro nombre, y nos los reparten según los vamos necesitando a lo largo de la vida. Cada vez que tienes que «echar un volado» o de alguna manera menos ostentosa tomar una decisión arriesgada, tu cerebro abre el sobre y saca el siguiente número «aleatorio», y permite que su valor *determine* lo que haces, tal como la lista de jugadas en «piedra, papel o tijeras». «Pero eso es injusto», alguien dirá, «porque a algunas personas les habrán repartido más números ganadores que a otras». En efecto, en cualquier reparto de las cartas específico, a algunas personas les tocarán más cartas altas que a otras, pero deberíamos recordar que, a la larga, la suerte se promedia. «¡Pero si todos los talones se sacan antes de que hayamos nacido, algunas personas están *determinadas* a tener más suerte que otras!» Sin embargo, eso sería cierto aunque los talones no se sacaran antes de nuestro nacimiento, sino periódicamente, sobre pedido, a lo largo de nuestras vidas. Incluso en una extracción perfectamente aleatoria e imparcial, una extracción genuinamente indeterminada, de todas maneras está *determinado* que algunas personas obtendrán más números ganadores que otras. Incluso en un torneo de volados perfectamente imparcial y justo, perfectamente aleatorio, está determinado que uno u otro ganará, y todos los demás participantes perderán. No sería correcto que el ganador afirmara que ganar era su «destino», pero cualesquiera ventajas asociadas con el triunfo son suyas, destino o no destino, ¿y qué podría ser más imparcial? La imparcialidad no consiste en que todo el mundo gane.

Acaso la razón más frecuentemente aducida para tener esperanzas de que exista el indeterminismo es que sin él, cuando elegimos un acto «no podríamos haberlo hecho de otro modo», y **no cabe duda** (¡ring!) de que eso es algo que debería importarnos. Tampoco esto es tan evidente como a menudo ha parecido, y para vislumbrar cómo podría esta idea familiar estarnos induciendo a error, consideremos la curiosa categoría de los hechos históricos inertes.

69. Hechos históricos inertes

Un hecho histórico inerte es cualquier hecho acerca de una disposición perfectamente ordinaria de la materia en el mundo del pasado que ya no es distinguible, un hecho del que no hay *absolutamente* ninguna huella en el mundo de hoy. Mi ejemplo favorito de un hecho histórico inerte es éste:

A. Algo del oro que tengo en los dientes perteneció alguna vez a Julio César. O a lo mejor este otro:

B. Es falso que algo del oro que tengo en los dientes haya pertenecido alguna vez a Julio César.

Ahora bien (nos dice la lógica), uno de éstos tiene que ser un hecho. (Espera un momento: ¿no era una moraleja del apartado sobre el operador maomeno que debíamos desconfiar de estas disyunciones «obvias»? Revisemos ésta: ¿cómo podría ser que ni A ni B tuvieran una clara prerrogativa con la verdad? Bueno, ¿qué tal si en tiempos de César la *propiedad* seguía siendo un concepto vago o mal definido, así que César sólo era maomeno dueño de parte de su oro…, quizá como la reina de Inglaterra hasta nuestros días es la dueña de todos los cisnes del reino?)[13] Suponiendo que el concepto pertinente de propiedad está bien demarcado, tal como lo está el concepto de oro, una de las dos oraciones debe expresar un hecho, pero cuál de las dos sea verdadera es algo que casi con toda seguridad no podría descubrirse con ninguna investigación física, por sofisticada que fuera y por mucho tiempo que tomara llevarla a cabo.

¿De verdad? Podemos imaginar ejemplos en los que podríamos estar casi seguros de que A, o bien B, es la alternativa verdadera. Si resultara ser que, gracias a diversos procesos históricos debidamente registrados, la «procedencia» de una parte del oro que tengo en los dientes se hubiera controlado y registrado escrupulosamente a lo largo de los milenios (como la «cadena de custodia» para exhibiciones de pruebas en los juicios por asesinato), podríamos estar completamente seguros de que A es verdadera. Digamos que mi dentista le compró a un museo un antiguo anillo de oro, el famoso «anillo para el dedo meñique del César», que una gran cantidad de documentos atestiguan que se había transmitido de monarca en monarca a lo largo de los siglos hasta que llegó al museo, y hay una videograbación que muestra al dentista fundiendo el anillo y vertiendo el oro fundido en el molde de yeso para mi tapadura. Descabellado, sin duda, pero claramente dentro de los límites de la posibilidad física. O en vez de eso, supongamos que soy un aficionado al bateo de oro y viajo a la base de un glaciar en retroceso en Alaska que había mantenido la tierra cubierta 10 000 años, donde yo escrupulosamente recolecté

todo el oro que después llegó a mis dientes. Entonces con todavía mayor seguridad B sería la verdadera. Pero si nada parecido a esos relatos extremos está bien fundado en las pruebas, entonces es prácticamente seguro que nunca podríamos saber cuál es verdadera, si A o B. La verdadera, sea cual sea, es un hecho histórico inerte.

La física cuántica plantea una dificultad interesante: incluso si pudiera seguírsele la pista a la trayectoria de cada átomo de oro en mis dientes a lo largo de los siglos, si hubiera habido ocasiones en que dos o más átomos de oro —uno del legado de César y otro de fuentes externas— colisionaron (o llegaron a estar *muy* cerca uno de otro), sería imposible en principio saber qué átomo era cuál después de la «colisión». Los átomos y las partículas más pequeñas no tienen nada parecido a huellas digitales u otras señas particulares, y no puede seguírseles la pista *continuamente,* así que la identidad continuada de partículas no siempre tiene sentido, lo que añade otra barrera al conocimiento de los hechos acerca del oro.

Ahora bien, sea o no determinista todo el universo, las computadoras están diseñadas para ser deterministas ante el ruido submicroscópico e incluso la aleatoriedad cuántica, y por ser digital y no análoga absorbe estas fluctuaciones (vimos un vívido ejemplo de eso en el apartado 66 con el **juego de la vida de Conway**, pero el determinismo digital está en todas partes). La idea fundamental detrás de la digitalización para producir determinismo es que podemos *crear* hechos históricos inertes mediante el diseño. Si clasificamos forzosamente todos los sucesos capitales en dos categorías —alto o bajo, encendido o apagado, 0 o 1— se garantiza que las microdiferencias (entre diferentes altos voltajes, diferentes maneras de estar encendido, diferentes tonos de 0) sean despiadadamente desechadas. No se permite que nada dependa de ellas y se desvanecen sin dejar rastro: hechos acerca de variaciones históricas reales que *dan absolutamente lo mismo* a las subsecuentes series de estados por los que pasa la computadora. Por ejemplo, tu amigo descarga de un sitio web una canción y la quema en dos CD. Estos dos CD sin etiquetar —los llamaremos A y B pero no escribiremos su nombre en ellos— serán, desde luego, duplicados digitales. Pídele que te «copie» uno de ellos en tu computadora portátil metido en un cuarto muy oscuro. No le digas qué CD usar, y tras haber completado la operación haz que coja los dos CD (para eliminar la posibilidad de que las huellas digitales, o rastros de ADN, se empleen como prueba), para después meterlos en una bolsa llena de otros CD y agitarla fuertemente. Ahora tenemos dos candidatos a hecho histórico inerte:

a) Tu computadora tiene una copia que se hizo con el disco A.
b) Tu computadora tiene una copia que se hizo con el disco B.

Cualquier codificación física real del flujo de bits de la canción tendrá una estructura

fina microscópica que difiere de la estructura fina de cualquier otra codificación. Y cuando «copies» uno de los CD a RAM (siglas inglesas de *memoria de acceso aleatorio*), los patrones de voltaje en RAM también tendrán una estructura fina única, y si luego «copiamos» el archivo de RAM a un disco duro o a una memoria USB, también éstos contendrán diferencias microscópicas que los distingan. Lo que convenientemente llamamos copiar es siempre la creación de otra señal física continua o análoga que tiene su propia estructura fina única porque así es como es el mundo entre los electrones y los protones. Pero el rasgo de la digitalización es que toda esta estructura fina se «ignore» y se borre en la «corrección de la norma». Cuando tienes normas, como un alfabeto, da lo mismo qUé *sAboR* tengan los símbolos individuales: todos se leen igual. En una computadora todo son 0 y 1.

Así, a menos que resulte que uno de los cd tiene un «error» distinguible en el nivel de la idealización digital (tiene un *bit invertido,* es decir, un 0 en vez de un 1, o viceversa), no habrá manera de saberlo. La digitalización impide la *propagación* de la individualidad de los dos cd a versiones posteriores, y finalmente a la conversión de digital a análogo que hace funcionar las bocinas o los audífonos. Algunos melómanos gozan fama de tener «oído de oro» y saber distinguir entre los discos de vinilo y los mejores cd digitales, o entre archivos digitales comprimidos (*e. g.,* mpeg) y archivos no comprimidos, pero a ningún melómano le puedes pedir que reconozca, si reproduces dos veces seguidas una misma pieza, si pusiste el mismo cd o dos distintos (y si adivina será por pura casualidad). Eso es un hecho histórico inerte, que no sólo es imperceptible para los oídos humanos, sino que ni siquiera viendo la estructura fina de sus «copias» en la memoria ram a través de un microscopio electrónico se puede detectar. Un ser humano que pudiera hacerlo sería sospechoso de tener percepción extrasensorial, pues en el proceso de digitalizar se crea una barrera física a la transmisión de la información que permitiría que alguien adivinara con algún fundamento y no sólo por azar.

Como una computadora es un aparato digital, no cuesta nada hacer que ejecute unos cuantos billones de pasos, y luego regrese *exactamente* al mismo estado (digital) que tenía antes, y verla ejecutar *exactamente* los mismos billones de pasos (digitales) una, otra y otra vez.

Espera un minuto, viene una objeción: ¿dices que las computadoras son deterministas? ¿Puedes hacer que repitan exactamente el mismo billón de pasos una y otra vez? ¡No me digas! ¿Entonces por qué mi computadora falla cada tanto? ¿Por qué mi procesador de palabras se congeló el martes mientras yo hacía exactamente lo mismo que el lunes, cuando funcionó perfectamente bien?

No estabas haciendo *exactamente* lo mismo. Si se congeló no es porque sea

indeterminista, sino porque el martes no estaba *exactamente* en el mismo estado que el lunes. Tu computadora debe de haber hecho en ese intervalo algo que levantó una «bandera» oculta o que llamó a alguna parte del procesador de palabras que nunca antes habías activado, que en algún lugar invirtió un bit que se guardó en su nueva posición cuando apagaste la máquina, y ahora el procesador de palabras se tropezó con ese cambiecito y falló. Y si tú de alguna manera consigues ponerla *exactamente* en su estado del martes en la mañana una segunda vez, volverá a fallar.

¿Y qué hay del «generador de números aleatorios»? Pensé que mi computadora tenía incorporado un dispositivo para crear aleatoriedad sobre pedido.

En estos días todas las computadoras traen incorporado un generador de «números aleatorios» que cualquier programa que se esté ejecutando en él puede consultar cuando lo necesite. (Antes de que hubiera computadoras podía comprarse un libro que no tenía más que una tabla de números aleatorios que podías usar en tu investigación: página tras página de dígitos que se habían generado escrupulosamente para pasar todas las pruebas de aleatoriedad que los matemáticos habían inventado. Desde luego, cada ejemplar de una edición determinada de esos libros tenía exactamente la misma secuencia de números aleatorios. rand Corporation publicó uno de los mejores en 1955. Constaba de un millón de dígitos aleatorios.) La secuencia de números que genera uno de los llamados generadores de números aleatorios no es en verdad aleatoria, sino sólo seudoaleatoria: es «matemáticamente compresible» en el sentido en que su infinitamente larga secuencia puede capturarse en un mecanismo finitamente especificado que luego la producirá de forma automática. Supongamos, por ejemplo, que tu generador de números aleatorios es un programa que puede especificarse en, digamos, un megabyte —ocho millones de bits — pero genera una secuencia (siempre la misma) que de hecho es infinita. Si quisieras mandarle a alguien esta serie infinita, no tendrías que hacerlo mediante un infinitamente largo mensaje de correo electrónico que registrara la serie *textualmente,* podrías tan sólo mandarle tu algoritmo del tamaño de un megabyte y él tendría acceso a la serie infinita completa. Ésta es la idea fundamental de un generador de números seudoaleatorios. Cada vez que arranques en frío el generador —cada vez que reinicies tu computadora, por ejemplo—, siempre arrojará *exactamente* la misma secuencia de dígitos, pero una secuencia que tiene la misma *apariencia* de falta de regularidad que si la hubieran generado fluctuaciones cuánticas genuinamente aleatorias. Es, podría decirse, una «tabla de números aleatorios» incorporada, un poco como lo que obtendrías en un bucle de cinta de video muy largo que grabara la historia de una ruleta a lo largo de millones de giros. El bucle siempre vuelve al «principio» cuando arrancas tu computadora. A veces esto importa. Los programas de computadora que aprovechan la aleatoriedad en diversos

puntos de «elección» repiten, no obstante, exactamente la misma secuencia de estados si arrancas en frío una y otra vez, y en ocasiones, si quieres probar si un programa tiene errores de programación, te saldrá siempre la misma «muestra aleatoria» de estados, a menos que hagas algo (esto es muy fácil) que le dé al programa un empujoncito que le permita de vez en cuando buscar su siguiente número «aleatorio» en algún otro lugar del flujo de dígitos.

70. Maratón de ajedrez por computadora

Es endiabladamente difícil pensar con claridad sobre el determinismo y la elección. Si el determinismo es verdadero, ¿alguna vez tenemos elecciones *reales*? Si un agente que aparentemente tiene libre albedrío es en realidad determinista, pues vive en un mundo determinista, ¿elimina esto toda elección, toda *oportunidad*? Te presento una bomba de intuición que para explorar esa pregunta mira un mundo simplificado (el juego de ajedrez) en un mundo determinista artificialmente construido: el mundo de la actividad que realiza una computadora.

Supongamos que instalas dos diferentes programas de ajedrez en tu computadora y los enyuntas con un programita supervisor que los enfrenta, juego tras juego, en una serie potencialmente interminable. ¿ Jugarán el mismo juego, una y otra vez, hasta que apagues la computadora? *Podrías* planearlo así, pero entonces no aprenderías nada interesante de los dos programas, A y B. Supongamos que A le gana a B en esta partida a menudo repetida. No podrías inferir a partir de eso que A fuera en general un mejor programa que B, o que A le ganaría a B en una partida diferente, y tampoco podrías aprender de la repetición exacta nada acerca de las virtudes y defectos de los dos programas. Sería mucho más informativo planear el torneo de modo que A y B jugaran una sucesión de partidas distintas. Esto se puede arreglar fácilmente. Si alguno de los programas consulta el generador de números aleatorios durante sus cálculos (si, por ejemplo, periódicamente «echa un volado» para librarse de situaciones donde no tiene ninguna razón práctica para hacer una cosa y no otra en el curso de su búsqueda de una buena jugada), entonces para la siguiente partida el estado del generador de números aleatorios habrá cambiado (a menos que dispongas que se reinicie), y por lo tanto se explorarán diferentes alternativas, en diferente orden, y eso hará que en ocasiones se «elijan» diferentes jugadas. Florecerá un juego variante, y el tercer juego será diferente de diferentes maneras, y esto dará por resultado una serie en la que ningún juego es igual a otro, tal como no hay dos copos de nieve iguales. Sin embargo, si apagaras la computadora y luego la reiniciaras ejecutando el mismo programa, se repetiría exactamente la misma serie abigarrada de partidas, porque la misma serie seudoaleatoria de números determinaría todos los «volados» usados por ambos programas.

Supongamos entonces que montamos un universo de ajedrez como éste en el que participan dos programas, A y B, y estudiamos los resultados de ejecutar, digamos, 1 000 partidas. Descubriremos montones de regularidades altamente confiables. Supongamos que descubrimos que A siempre le gana a B en 1 000 partidas *diferentes*. Ésa es una regularidad que querremos explicar, y algo como decir «Ya que el programa es determinista, se *causó* que A siempre le ganara a B» difícilmente conseguiría responder a nuestra muy razonable curiosidad. Querremos saber qué hay en la estructura, los métodos, las disposiciones de A, que explica su superioridad en el ajedrez. A tiene una

competencia o un poder que a B le falta, y necesitamos aislar este interesante factor. Podría ser que la explicación yaciera en un bajo nivel; podría resultar, por ejemplo, que el programa A y el programa B fueran de hecho el mismo programa, evaluadores de jugadas de ajedrez *idénticos* en el nivel del código fuente, pero que el programa A fuera un compilador más eficiente que B y por eso pudiera explorar más a fondo la partida que el programa B en el mismo número de ciclos máquina. En efecto, A «tiene exactamente los mismos pensamientos» sobre el ajedrez que B, y B «sabe» de ajedrez lo mismo que A, sólo que A piensa más rápido. (El ajedrez serio, el de los torneos, siempre se juega con un reloj; si se te acaba el tiempo antes de haber hecho todas tus jugadas, pierdes.) Es más probable que la superioridad de A con respecto a B requiriera explicación en una perspectiva de más alto nivel en la que aparecieran los asuntos cotidianos de toma de decisiones del ajedrez: *representaciones* de posiciones en el tablero, *evaluaciones* de continuaciones posibles, *decisiones* sobre qué continuaciones explorar más a fondo, etc. Así, puede ser que el programa A ajuste el valor relativo de sus piezas conforme avanza la jugada, o que tenga mejores funciones evaluadoras de las posiciones del tablero, o que decida terminar antes o después ciertas clases de exploraciones. No «tiene los mismos pensamientos» que B, sino que «tiene mejores y más sofisticados pensamientos» (**maomeno** tiene estos pensamientos, desde luego: no es una persona consciente).

De hecho, podría ser más revelador que un programa no siempre ganara. Supongamos que A *casi siempre* le gana a B, y supongamos que A evalúa las jugadas mediante un diferente sistema de principios. Entonces tendríamos algo incluso más interesante que explicar. Para investigar *esta* cuestión causal tendríamos que estudiar la historia de las 1 000 partidas distintas, en busca de futuras regularidades. Seguro que encontraríamos muchas. Algunas serían endémicas del ajedrez se juegue donde se juegue (*e. g.,* la casi certeza de que B va a perder en cualquier juego donde B deje rezagada a una torre), mientras que algunas serían peculiares de A y B en tanto que son jugadores de ajedrez específicos (*e. g.,* la tendencia de B a sacar demasiado pronto a la reina). Encontraríamos las regularidades estándar de la estrategia del ajedrez, como el hecho de que cuando se le está acabando el tiempo a B, busca menos a fondo en los nodos restantes del árbol de la partida que cuando le queda más tiempo y está en la misma ubicación. En pocas palabras, encontraríamos una abundancia de regularidades *explicativas,* algunas sin excepciones (en nuestra ejecución de 1 000 partidas) y otras estadísticas.

Esas regularidades que se pueden reconocer en las jugadas del ajedrez son momentos destacados en el desarrollo de un espectáculo determinista que, observado desde la perspectiva de la microcausalidad, es prácticamente igual. Lo que desde una posición estratégica nos parece que son dos programas de ajedrez en un combate lleno de suspenso, a través del «microscopio» (tal como vemos las instrucciones y los datos fluir

desde el CPU de la computadora) puede verse como un autómata determinista desarrollándose del único modo que puede, con sus brincos que ya pueden predecirse si se examina el estado preciso del generador de números seudoaleatorios y el resto del programa y de los datos. No hay bifurcaciones o ramas «reales» en su futuro; todas las «elecciones» que hacen A y B ya están determinadas por el estado total de la computadora y su memoria. Parece que nada es realmente *posible* en este mundo salvo lo que realmente pasa. Supón, por ejemplo, que una ominosa *red de mate* (una victoria garantizada que podría ser difícil de reconocer) amenaza a A en el tiempo *t,* pero se colapsa cuando se le acaba el tiempo a B y termina su búsqueda de la jugada clave un pulso antes de tiempo. Esa red de mate *nunca iba a ocurrir.* (Si lo dudáramos, esto es algo que podríamos demostrar ejecutando exactamente el mismo torneo otro día. En el mismo momento de la serie, a B otra vez se le acabaría el tiempo y terminaría su búsqueda exactamente en el mismo punto.)

¿Qué debemos decir? ¿En verdad este mundo de juguete es un mundo en el que nada se puede prevenir ni evitar, no hay ofensiva y defensiva, no hay oportunidades perdidas ni posibilidades genuinas, donde falta el estira y afloja de la genuina acción? Hay que reconocer que nuestros programas de ajedrez, como los insectos y los peces, son agentes demasiado simples para ser candidatos viables a tener un libre albedrío moralmente significativo, pero el determinismo de su mundo no los priva de sus diferentes poderes, sus diferentes *habilidades* para aprovechar las *oportunidades* que se les presentan. Si queremos entender qué está pasando en ese mundo, podemos —de hecho debemos— hablar acerca de cómo sus *decisiones* informadas hacen que sus circunstancias cambien, y acerca de lo que *pueden* y *no pueden* hacer. Si queremos dejar al descubierto las *regularidades causales* que explican las regularidades y similitudes que hallamos en esas 1 000 partidas, tenemos que tomarnos en serio la perspectiva que describe el mundo como uno que contiene a dos agentes, A y B, que tratan de vencer al otro en el ajedrez.

Supongamos que manipulamos el programa para que cada vez que gane A suene una campana y cada vez que gane B suene una chicharra. Empezamos el maratón y una observadora que no sabe nada del programa nota que la campana suena muy a menudo y la chicharra casi nunca. Ella quiere saber qué explica esta regularidad. Puede reconocerse la regularidad con la que A le gana a B y *describirse,* independientemente de que se adopte la **postura intencional**, pero sigue haciendo falta una *explicación.* La única explicación —la explicación correcta— puede ser que A genera mejores «creencias» sobre lo que B hará si… que las que genera B sobre lo que A hará si… En tal caso, se *necesita* adoptar la postura intencional para encontrar la explicación (véanse los apartados 33 y 42 para otros ejemplos de conexiones causales que son completamente inexplicables a menos que adoptes la postura intencional).

Hasta ahora bien, pero estas «decisiones» y «elecciones» parecen no ser más que

maomeno decisiones y elecciones. Parece ser que les falta algo que sí tienen las elecciones genuinas: «podría haberlo hecho de otro modo». Pero miremos más de cerca un ejemplo específico, pues las apariencias engañan. Para esto ayudará añadir un tercer programa de ajedrez, C, a nuestro programa de torneo. Supongamos que C es mejor que A y que B, y que casi todo el tiempo les gana a los dos, y supongamos también que las primeras 12 jugadas en dos de estas partidas son exactamente las mismas, y que C gana ambas partidas, derrotando tanto a A como a B, aunque por caminos un poco diferentes después de esos primeros 12 movimientos. Los expertos se reúnen para un análisis retrospectivo y se dan cuenta de que en la jugada 12, la última jugada común, si A o B hubieran enrocado, C probablemente habría perdido. Enrocar en la jugada 12 era la llave a la victoria, y tanto a A como a B se les escapó.

La diseñadora del programa se encoge de hombros y dice: «Bueno, A podría haber enrocado», y la diseñadora de B añade: «Mi programa también; B podría haber enrocado». ¡*Pero la diseñadora del programa A tenía razón y la diseñadora del programa B no!* ¿Cómo es posible? El programa de torneos T es determinista, y si ejecutamos todas las partidas otra vez, *exactamente* en el mismo estado, ni A ni B enrocan. ¿No se estará engañando la diseñadora del programa A? No necesariamente. ¿Qué es lo que tratamos de descubrir cuando preguntamos si A podría haberlo hecho de otro modo? Mirar *precisamente* el mismo caso, una y otra vez, no aporta ninguna información, pero mirar casos *semejantes* permite, de hecho, obtener un diagnóstico. Si descubrimos que en muchas circunstancias semejantes en otras partidas A *sí* continúa un poco más con la evaluación, y descubre las virtudes de esas jugadas y las realiza, entonces apoyamos la convicción de la diseñadora de que A podría haber enrocado en ese momento.

En el ejemplo mínimo, podríamos descubrir que invertir un solo bit en el generador de números (seudo)aleatorios habría dado por resultado que A enrocara. Supongamos que la diseñadora de A cava más hondo en la ejecución actual del programa y muestra que en esta ocasión A dejó de «pensar» un pulso antes de tiempo (todos los programas de ajedrez, por brillantes que sean, tienen que truncar arbitrariamente sus búsquedas en algún punto). A consideró la posibilidad de enrocar, y había empezado a analizar ese resultado, pero como se le acababa el tiempo, consultó su generador de números aleatorios echando, en efecto, un volado, y se decidió por la jugada que hasta ese momento había identificado como la mejor (y no era el enroque). Pero si el número seudoaleatorio hubiera sido un 1 en vez de un 0, A habría seguido considerando por un poco más de tiempo la posibilidad de enrocar, y lo habría hecho. «¡Sólo invierte un bit en el número aleatorio y A gana!», dice la diseñadora. Diríamos que, en este caso, el hecho de que A no haya enrocado fue una casualidad, mala suerte con el generador de números aleatorios.

Cuando volteamos a ver a la diseñadora de B, no cabe esperar una historia así que pudiera apoyar la afirmación de que B habría podido enrocar bajo esas circunstancias. Es cierto que B «sabe» que enrocar es una jugada legal en estas condiciones, y quizá brevemente «consideró» la posibilidad de enrocar, pero B no estuvo cerca, ni por asomo, de elegir el enroque en esta ocasión. El enroque era una jugada *profunda,* la clase de jugada que en las columnas de ajedrez de los periódicos viene mencionada con signos de admiración, y está mucho más allá de los limitados poderes analíticos del programa B. Así, tenemos aquí un mundo completamente determinista —el programa T— en el que A podría haber enrocado pero B no podría haber enrocado. La diferencia entre A y B es real y explicativa: es una diferencia en competencia o habilidad. Eso podría expresarse de una manera aparentemente paradójica:

> A podría haber enrocado en el momento *t,* pero el universo no podría haber tenido un suceso de enroque en el momento *t.*

¿Qué podría autorizarnos a describir así la situación? Sencillamente esto: si consideramos a A *separada de su medio inmediato —que incluye el generador de números aleatorios—,* es indeterminado si A enroca o no. Depende de algo que estrictamente hablando está afuera de A. Dada la manera como está ordenado el resto del universo en *t,* el enroque no era posible para A, pero eso «no es culpa de A». Por el contrario, B no podría haber enrocado: no estaba en su naturaleza. Para imaginar a B enrocando se necesitarían demasiadas alteraciones de la realidad.

Esto es un descubrimiento útil: una distinción entre lo que A y B «podrían» hacer *no* depende del indeterminismo. Incluso en un mundo determinista podemos ver que A *puede hacer* tipos de cosas que B *no puede hacer,* y esta diferencia es parte de la *explicación* de por qué A le gana a B. Sencillamente, no es interesante el hecho de que, porque el determinismo es verdadero en este mundo, A y B sólo puedan hacer lo que de hecho hacen en la ocasión específica (y harían una y otra vez si se repitieran *exactamente* las mismas circunstancias); esto es irrelevante para la explicación que obtenemos de la regularidad perfectamente objetiva y visible: A le gana a B.

Un programa de ajedrez no es un agente moral, y no es moralmente responsable de las decisiones que toma —su mundo es completamente amoral; transgredir alguna de las reglas del ajedrez es impensable para un programa de ajedrez, y por lo tanto no necesita castigos para las transgresiones—. Sin embargo, como acabamos de ver, hasta en el simple mundo determinista del ajedrez por computadora podemos hacer una distinción importante y real entre A y B. A veces, cuando A hace alguna tontería o algo inteligente podemos decir: «A podría haberlo hecho de otro modo, pero B no podría haberlo hecho de otro modo». Si piensas que esto ha de ser un error «porque ni A ni B podrían haberlo

hecho *jamás* de otro modo pues el mundo es determinista», eres tú quien comete el error.

A y B difieren en sus competencias ajedrecísticas, y «podría haberlo hecho de otro modo» recoge bien un aspecto de esa diferencia, como acabamos de ver. ¿Y qué decir de la competencia *moral*? Cuando la gente, al hablar de unos seres humanos que actúan mal, que «podrían haberlo hecho de otro modo», y usa esto como su justificación para no disculparlos, y al mismo tiempo concuerda en que otros seres humanos en circunstancias similares *no* podrían haberlo hecho de otro modo, tampoco está cometiendo un error, y esto es independiente de si el determinismo es verdadero o no. Están apuntando a una *diferencia real en competencia moral* que no depende ni del indeterminismo ni del determinismo, y que puede cimentar una diferencia en nuestra respuesta.

Para verlo más claramente, adopta la perspectiva de la programadora que diseñó el programa B. Ella quiere saber si ha descubierto una debilidad de B. Tenemos aquí una partida en la que no enrocar le costó a B la victoria; ¿podría haber enrocado entonces? Si todo lo que se necesitara para que eso ocurriera fuera la inversión de un solo bit en el generador de números aleatorios, entonces quizá no hace falta ninguna mejora de diseño. La mitad de las veces, en circunstancias similares, B enrocará, y quizá eso es lo máximo a lo que cualquiera podría aspirar. Un programa siempre debe usar números aleatorios de vez en cuando (como volados) para terminar una búsqueda y seguir adelante con el juego, y por lo tanto siempre habrá casos en los que gracias al volado la búsqueda se detenga justo antes del descubrimiento. Y observa que la situación no mejora si le damos al programa B (o al programa A) un generador de números aleatorios cuántico, digamos, un contador Geiger que arroja bits basados en las trayectorias indeterminadas de las partículas subatómicas. Y ahora piensa en lo que podríamos decir de B en el caso de que B no enrocara debido a un solo 0 donde debería haber un 1. Si el generador de números cuántico arroja un 0, B enroca; si arroja un 1, B no enroca. «B podría haber enrocado», dice un observador cuando sale el 1. Sí, pero B no por eso es más libre. En una serie de partidas en las que surge esta clase de oportunidad, la mitad de las veces B enrocará y la mitad de las veces no, sin importar si el generador de números aleatorios de B es «genuino» o «seudo». El filósofo David Wiggins escribió sobre la «injusticia cósmica» del determinismo, pero nuestra bomba de intuición sobre el torneo de ajedrez por computadora muestra que la «injusticia cósmica» del indeterminismo es igual.[14] B está «a merced» de su generador de números aleatorios *o* de su generador de números seudoaleatorios (también A, claro está; también todos nosotros). No hay razón para preferir el generador de números genuinamente aleatorios, a menos, claro, que estés planeando jugar al ajedrez contra un Dios omnisciente que pueda examinar tu generador de números seudoaleatorios y planear sus jugadas acorde con eso.

Así, seguimos buscando una razón para querer que el indeterminismo sea verdadero. A lo mejor podemos tener todo el libre albedrío que valga la pena desear sin que el indeterminismo tenga ningún papel. Aquí hay otra aspirante a razón:

> No puedo cambiar el pasado, pero si el indeterminismo es verdadero, ¡puedo cambiar el futuro!

No, no. Cambiar el futuro ¿de qué a qué? ¿De lo que *iba* a ser a lo que *va* a ser? Tan no puedes *cambiar el futuro* como no puedes *cambiar el pasado*. El concepto es incoherente. Entonces:

> Si el determinismo es verdadero, no puedo cambiar el futuro, y si el determinismo es falso, no puedo cambiar el futuro. Se sigue entonces que no puedo cambiar el futuro.

¿Por qué parece que queremos cambiar el futuro? Porque queremos poder prever los desastres y hacer algo para que no ocurran. Y *podemos* hacerlo, independientemente del indeterminismo. Si alguien te avienta un ladrillo, y tú lo ves y lo esquivas, puedes evitar que el ladrillo te pegue. ¡Bien hecho! ¿Iba a ocurrir la colisión? En un sentido sí, porque está claro que la trayectoria del ladrillo iba directo a tu cabeza, pero como lo viste (porque la luz que rebotaba de él a tus ojos hizo que tu cerebro calculara el riesgo y tomara alguna medida), lo evitaste. Desde luego, si hubieras querido *evitar evitarlo* (si se te hubiera ocurrido alguna razón por la que de hecho estaba mejor que dejaras que te golpeara), podrías haberlo hecho. Un observador no podría saber, hasta el último momento, si ibas a recibir el golpe o no. Y si hubiera apostado a que lo esquivarías, perdería. Estamos de vuelta en nuestra razón para querer ser impredecibles, para lo cual no hace falta el indeterminismo.

¿Qué consigue esta bomba de intuición? Toma la conocida frase «podría haberlo hecho de otro modo» y muestra que, en contra de la generalizada pero mal estudiada opinión, una versión valiosa de ella *no* depende del indeterminismo. Si hay un sentido de «podría haberlo hecho de otro modo» que sea al mismo tiempo incompatible con el determinismo y moralmente importante —no sólo una curiosidad metafísica, podría decirse—, tiene que demostrarse, y la carga de la prueba está del lado de los que lo creen. Otro tema «obvio» que desenmascaramos como a fin de cuentas no tan obvio.

71. Máxima responsabilidad

Hasta ahora hemos visto decisiones triviales que no suponen ninguna responsabilidad moral: piedra, papel o tijeras, jugadas de ajedrez y ladrillos esquivados. A lo mejor cuando veamos específicamente nuestros intentos de ser agentes *morales,* no sólo sistemas intencionales como las computadoras que juegan al ajedrez y las gacelas que brincan, el indeterminismo será verdaderamente deseable. Muchos pensadores así lo han creído. Para ellos, estos ejercicios no son más que una distracción. Te presento una versión muy clara de lo que algunos pensadores consideran el argumento decisivo. Lo planteó de esta forma el filósofo Galen Strawson:[15]

1. En cualquier situación dada, haces lo que haces debido a tu forma de ser.
2. Así, a fin de ser máximamente responsable de lo que haces, tienes que ser máximamente responsable de tu forma de ser, al menos en ciertos aspectos mentales de primordial importancia.
3. Sin embargo, en ningún aspecto puedes ser máximamente responsable de tu forma de ser.
4. Entonces no puedes ser máximamente responsable de lo que haces.

La primera premisa es innegable: «tu forma de ser» pretende incluir tu estado total en el momento, sin importar cómo llegaste a él. Sea el estado que sea, tu acción fluye de él de una manera no milagrosa. La segunda premisa señala que no podrías ser «máximamente» responsable de lo que haces salvo si fueras «máximamente» responsable de haber llegado a ese estado, al menos en unos aspectos. Sin embargo, según el paso *3* esto es imposible.

El paso *4,* la conclusión, sí parece seguirse lógicamente. A varios filósofos este argumento les ha parecido decisivo e importante. ¿Pero de verdad lo es? Miremos más de cerca el paso *3*. ¿Por qué no puedes ser máximamente responsable de tu manera de ser, al menos en *algunos* aspectos? En la vida cotidiana hacemos exactamente esta distinción, e importa moralmente. Imagina que diseñas y construyes un robot y lo mandas al mundo solo, sin que nadie lo cuide, y sabiendo perfectamente bien a qué clase de actividades podría dedicarse, e imagina que daña seriamente a alguien. ¿No eres responsable de esto, al menos en algunos aspectos? La mayoría de la gente diría que sí. Tú lo construiste: tú debiste prever los peligros —de hecho *sí* previste algunos de los peligros—, y ahora tienes la culpa, al menos en parte, del daño ocasionado. Muy pocas personas sentirían alguna compasión si te empeñaras en que no eres responsable *en absoluto* del daño que ocasionó tu robot.

Ahora considera un ejemplo un poco diferente: diseñas y construyes a una persona (tú mismo en un momento posterior) y te mandas al riesgoso mundo sabiendo

perfectamente bien los posibles peligros a los que te enfrentarías. Te emborrachas en un bar y luego subes a tu coche y te vas manejando. ¿No eres responsable, al menos en parte, de «tu forma de ser» en el momento en que chocaste contra un camión escolar? De acuerdo con el sentido común, desde luego que sí (el camarero, o el servicial anfitrión, que te dejó subirte al coche, pueden compartir la responsabilidad). ¿Pero cómo sería posible en vista del argumento fulminante de Strawson? Bueno, recuerda que Strawson dice que no puedes ser *absolutamente* responsable de tu forma de ser. Está bien, ¿y eso qué? ¿A quién le parecería importante que fueras *absolutamente* responsable? Eso sí que es un estado completamente imposible, ¡*aunque el indeterminismo fuera verdadero!* (¡Y luego dicen que este argumento nos da una razón para desear el indeterminismo!) Esto es lo que dice Strawson: «Para ser absolutamente responsable de lo que uno hace, uno tendría que ser *causa sui,* la causa de uno mismo, y esto es imposible (ciertamente no sería más posible si tuviéramos almas inmateriales en vez de ser completamente materiales)».[16] La responsabilidad *absoluta* es una pista falsa, una bendición que nadie debería anhelar. Strawson no piensa lo mismo, y me critica por no darle el reconocimiento que merece: «Dennett no demuestra la clase de libre albedrío absoluto y de responsabilidad moral en los que la mayoría de la gente quiere creer y en los que en efecto cree. Eso no puede hacerse, y lo sabe».[17] Tiene toda la razón: no demuestro la clase de libre albedrío en que la mayoría de la gente quiere creer, lo sé bien. Sin embargo, creo que la mayoría de la gente *se equivoca* al querer creer en él, y si es que cree en él, también en esto se equivoca. Les toca a Strawson y a otros demostrar *por qué* debería interesarnos la responsabilidad máxima —o el problema del determinismo/indeterminismo— en nuestras vidas. Pueden definir una variedad de libre albedrío que sea incompatible con el determinismo, y mostrar que un montón de gente piensa que es importante, pero también tienen que demostrar que esa gente no se engaña. ¿Por qué debería interesarle a nadie? (Observa mi pregunta retórica. Me estoy exponiendo. Me encantaría que Strawson o alguien más aprovecharan la coyuntura y trataran de responderla, pero hasta ahora no hay ningún voluntario.)

Antes de dejar el argumento de Strawson, déjame preguntar si has notado su asombroso parecido con un argumento anterior. Lo expresaré de una forma un poco cercana a la de Strawson para destacar las similitudes.

1. En cualquier contexto dado, un mamífero es un mamífero debido a su forma de ser.
2. A fin de ser un mamífero, tienes que adquirir tu forma de ser teniendo de madre a un mamífero.
3. Sin embargo, esto debe ser verdadero de tu madre, y de su madre, y así sucesivamente *ad infinitum,* lo cual es imposible.
4. Entonces no puedes ser un mamífero porque los mamíferos son

completamente imposibles.

Siempre debes recelar de los argumentos «ancestrales» como éste. Es casi seguro que se trate de un ejemplo disfrazado de la antigua falacia conocida como el argumento *sorites* (o «paradoja del montón»):

1. Un solo grano de trigo no es un montón.
2. Agregarle un grano a un solo grano no lo convierte en montón.
3. No puedes convertir un no montón en montón por agregarle un solo grano.
4. ¡Por lo tanto no existen los montones!

Los filósofos han escrito sobre la paradoja sorites y los problemas con los límites vagos de los términos (que a todas luces es de lo que depende la paradoja) desde hace miles de años, y todavía no hay un acuerdo sobre exactamente cómo diagnosticar y evitar la falacia. (Véase la *Stanford Encyclopedia of Philosophy* en línea para una magnífica y actualizada investigación.) Incluso hay unos cuantos filósofos valientes que declaran válido el sorites, y tratan de vivir con «el hecho» de que no hay hombres calvos y tampoco hombres no calvos, una posición muy difícil de defender. Pero como muestra el apartado 43, Darwin nos enseñó cómo darle la espalda al sorites: no necesitamos encontrar líneas divisorias «con reglas claras» clasificadas en el orden de los ancestros.

Hasta donde sé, soy el primero en señalar la semejanza del argumento de Strawson —y otros argumentos de la misma ralea en la bibliografía sobre el libre albedrío— con la paradoja sorites, pero helo ahí. Creo que es tan evidente que la gente puede gradualmente volverse moralmente responsable a lo largo de su paso de la infancia a la adultez, como que los linajes de reptiles y luego terápsidos pueden gradualmente convertirse en un linaje de mamíferos a lo largo de millones de años. No tienes que ser un mamífero *absoluto* para ser mamífero, y no tienes que ser *absolutamente responsable* para ser responsable, ni tener un libre albedrío *absoluto* para tener una especie de libre albedrío que valga la pena desear. De hecho, como un libre albedrío absoluto sería milagroso, se necesita un argumento muy convincente para mostrar por qué alguien codiciaría una cosa así. ¿Quieren ser Dios? Qué pena, hoy no están de suerte; pero ser lo mejor que hay después de eso ya es bastante bueno.

72. Sphexismo

Doug Hofstadter acuñó el término «sphexismo» para referirse a una clase conocida de rígido automatismo robótico que a menudo se confunde con gran inteligencia.[18] El ejemplo que define el sphexismo, y la inspiración del término, es una avispa con un curioso comportamiento. Doug y yo nos habíamos topado, cada uno por su lado, con un pasaje en un libro de divulgación científica, *The Machinery of the Brain,* escrito por Dean Wooldridge, quien describió así a la avispa *Sphex:*

> Cuando llega el momento de poner los huevos, la avispa *Sphex* cava una madriguera para ese fin y busca un grillo, al que pica de tal modo que lo paraliza pero no lo mata. Jala el grillo a la madriguera, a su lado pone los huevos, tapa la madriguera y vuela lejos, para nunca volver. A su debido tiempo, se rompen los huevos y las larvas de avispa se alimentan del grillo paralizado, que no se ha descompuesto gracias a que está en el equivalente de un congelador para una avispa. Para la mente humana, una rutina tan elaboradamente organizada y que parece tener un sentido comunica un convincente sabor de lógica y consideración… hasta que se analizan otros detalles. Por ejemplo, la rutina de la avispa es traer al grillo paralizado a la madriguera, dejarlo a la entrada, ir adentro para ver que todo esté bien, salir de ahí y luego meter al grillo. Si se mueve al grillo unos centímetros mientras la avispa está adentro haciendo su inspección preliminar, al salir de la madriguera devolverá al grillo a la entrada, pero no lo meterá, y entonces repetirá los preparativos de entrar a la madriguera a ver si todo está bien. Si otra vez se mueve al grillo unos centímetros mientras la avispa está adentro, de nuevo acercará al grillo a la entrada y volverá a entrar a la madriguera para una revisión final. A la avispa nunca se le ocurre meter al grillo sin más. En una ocasión el procedimiento se repitió 40 veces, siempre con el mismo resultado.[19]

Parece un ejemplo perfecto de la competencia imperfecta, carente de comprensión, que desenmascaramos cuando exponemos el seudoentendimiento superficial de un programa informático mediocre. Sin embargo, hace poco supimos que Wooldridge —como a menudo hacen los autores de divulgación científica— nos dio un bosquejo sobresimplificado del fenómeno. El psicólogo Lars Chittka me escribió una nota con una cita de la obra de Jean-Henri Fabre (de 1879), que por lo visto había sido la fuente de Wooldridge, quien, si hubiera seguido leyendo a Fabre, habría descubierto que de hecho sólo algunas avispas *Sphex* son sphexistas. De hecho, Fabre estaba muy deseoso de señalar este detalle. Si en un primer vistazo pensaste que la *Sphex* era lista, y en un segundo vistazo pensaste que la *Sphex* era estúpida, trata una tercera vez y verás que algunas avispas *Sphex* no son tan sphexistas a fin de cuentas. El original francés, que me

consiguió Roger DePledge, incluye la siguiente oración: «Après deux ou trois épreuves […] le saisit avec les mandibules par les antennes et l'entraîne immédiatement dans le terrier. Qui fut sot?» La versión que me mandó Chittka es la traducción de Fabre al alemán, que dice así: «Nach zwei oder drei Malen, […] packt ihre Fuehler MIT den Kieferzangen und schleift sie in die Höhle. Wer war nun der Dummkopf ?» («Después de dos o tres intentos […] tomó [a su presa] de las antenas con las mandíbulas e inmediatamente la metió al hoyo. ¿Quién es el tonto?»)

Así, pues, el adjetivo «sphexista» es un nombre poco apropiado, pero ya que ha tenido tanta aceptación, las avispas *Sphex* tendrán que aguantarse el insulto. De alguna manera tienen suerte por haber acaparado la atención del público general, lo que probablemente represente un no desdeñable aumento de aptitud. (¿Por la protección de qué hábitat votarías: el de la *Sphex* o el del rutinario escarabajo?) Las avispas *Sphex* podrán no ser «carismáticas» como los elefantes, los tigres y los lobos, pero son bastante conocidas —dado su problemático sphexismo—.

El sphexismo es una propiedad importante no tanto porque lo presenten tantos animales simples —insectos, gusanos, peces— (aunque sí lo hacen, en diferente grado) sino porque nos da un término para las competencias limitadas, robóticas, miopes, a partir de las cuales podemos construir mentes más elaboradas, versátiles, comprensivas. A los bloques de construcción en cualquier modelo de la mente ¡les convendría ser sphexistas! O, como observé antes, los bloques de construcción deberían ser maomeno mentes, pálidas sombras de nuestras mentes. El sphexismo también es útil para distinguir entre las mentes que tienen competencia moral y las que no. En la medida en que un ser humano sea sphexista (debido a un tumor o alguna lesión en el cerebro, a un grave desequilibrio de los neuromoduladores, por enfermedad mental o por simple y llana ignorancia o inmadurez), ese ser humano no podría haberlo hecho de otro modo en el sentido que viene al caso.

El biólogo perseverante que interfiere con la avispa *Sphex* para exhibir su sphexismo es el mismo modelo de agente manipulador al que con toda razón tememos. Muchos experimentos mentales de filósofos acerca del libre albedrío dependen de que se invoque un marionetista justo como ése, o una neurocirujana nefanda que a escondidas le ha cruzado a alguien los cables para que haga lo que a ella se le antoje. Me imagino que la moraleja de estos relatos de miedo es que, aunque no haya un marionetista de verdad, el hecho de que nuestra conducta esté causada por diversos rasgos de nuestros medios, procesados a través de nuestros sistemas perceptuales y nuestros cerebros, muestra que bien podría haber uno. (La ilustración de portada del librito de Sam Harris *Free Will* es el juego de los hilos con los que se mueve a las marionetas.)[20] Sin embargo, la conclusión a la que se invita es un claro *non sequitur*. Cuando el «control» que detenta el medio pasa por nuestros sistemas perceptuales que funcionan a la perfección y por nuestros no

engañados cerebros, no hay nada que temer. De hecho, nada es más deseable que las cosas y sucesos que nos rodean nos hagan tener creencias verdaderas acerca de ellos que luego podamos usar para modular nuestra conducta en nuestro provecho. Los fotones que rebotan de los respiraderos que quedan en la arena al retirarse la marea y llegan a mis ojos pueden hacerme agarrar mi rastrillo para almejas, mi canasta, y empezar a cavar. Si éste es un ejemplo de estar controlado por mi medio, estoy completamente de acuerdo. Y, como la mayoría de la gente, no me siento amenazado o manipulado cuando mis amigos me ofrecen cenas magníficas, sabedores de que seré incapaz de resistir la tentación de comerlas.

Algo más que notarás acerca de los ejemplos de marionetistas y neurocirujanos en lo que se ha publicado sobre el libre albedrío es que la intervención siempre —siempre— es secreta. ¿Por qué tendría que ser así? Porque sólo cuando alguien más, un agente secreto, nos hace actuar o elegir *contra nuestra voluntad* llueven las intuiciones de que nuestro albedrío o nuestra voluntad no son libres. La razón para esto es obvia y tiene su origen en la idea que inauguró la teoría de juegos: cuando un agente sabe del intento de manipulación de otro agente, busca contramedidas, y como mínimo ajusta su comportamiento para sobrellevar mejor este descubrimiento. Las interacciones competitivas entre los dos agentes suponen múltiples niveles de realimentación, y por lo tanto un disminuido «control» por el aspirante a manipulador. Y si la intervención no sólo no es secreta, sino que el «títere» o «marioneta» la solicita, la cosa se pone completamente de cabeza.

Podemos demostrarlo con sólo alterar un poco los ejemplos más comunes, girando las perillas de las bombas de intuición. El filósofo Harry Frankfurt inventó una bomba de intuición en la que una persona está siendo observada mediante aparatos de neuroestimulación que secretamente le implantaron, y los maneja un neurocirujano que se asegura de que esa persona decida lo que él quiere.[21] Si tú, como esa persona, te enfrentas a un dilema y eliges la opción A en vez de la opción B, entonces, si es lo que el neurocirujano quería que eligieras, no hace nada; si en lugar de eso sus aparatos muestran que estás a punto de elegir la opción B, aprieta el botón y lo impide, así que después de todo eliges la opción A. Tú no sientes nada. ¿En alguno de los dos casos elegiste libremente? A lo largo de los años los filósofos han escrito cientos de páginas acerca de «ejemplos de Frankfurt» con muchas variantes, pero hay un giro de las perillas que marca una enorme diferencia y que nunca se ha explorado, que yo sepa. Helo aquí:

Quieres dominar tu amor incontrolable a los postres engordadores, así que has contratado al buen doctor para que te implante ese dispositivo, y le pagas una generosa suma para que observe todas tus comidas, con lo que te proporciona una red de seguridad que te impide pedir helados con jarabe de chocolate y rebanadas de pastel de queso. Los dos esperan que él nunca tenga que apretar su botón, y pronto a ti

prácticamente se te olvida que él o su asistente están electrónicamente al alcance de tu mano. Te has curado y el aparato ha registrado una cadena ininterrumpida de cientos de elecciones A («No, gracias, sólo una taza de café sin azúcar»). Si eso no son ejemplos en los que *tú* decides responsable y libremente, ¿por qué no lo son? ¿No fue sensato de tu parte obtener ayuda para tu fuerza de voluntad un tanto frágil?

73. Los niños del Brasil: otra muleta de razonamiento

En 2004, los psicólogos Joshua Greene y Jonathan Cohen escribieron «For the Law, Neuroscience Changes Everything and Nothing», que se publicó en la prestigiosa revista *Philosophical Transactions of the Royal Society.* Este influyente artículo evoca una revolución en la ley desencadenada por descubrimientos científicos.

> La ley *dice* que no presupone nada más que una noción metafísicamente modesta del libre albedrío que es perfectamente compatible con el determinismo. Nosotros sostenemos, sin embargo, que el apoyo intuitivo de la ley está basado en última instancia en una noción libertaria [indeterminista] y metafísicamente demasiado ambiciosa del libre albedrío que se encuentra amenazada por el determinismo y, más explícitamente, por la neurociencia cognitiva que viene en camino.[22]

La conclusión a la que buscan llegar es sutil; tiene que serlo, pues conceden que hay una buena cantidad de argumentos que apoyan el compatibilismo (la postura que yo defiendo aquí) pero quieren mostrar que en realidad «todos estamos indecisos» con respecto al libre albedrío. Presentan un experimento mental ideado para revelar que el pensamiento de sentido común *cotidiano* depende mucho del determinismo. Observa que esto es muy parecido a lo que yo he estado pidiendo: muéstrennos por favor *lo que* debería preocuparnos sobre el indeterminismo y dígannos *por qué.* Su experimento mental se inspira en la película *Los niños del Brasil,* acerca de unos científicos nazis que crían unos clones de Hitler (gracias a un poco de ADN que se rescató). Helo aquí: «Supongamos, entonces, que un grupo de científicos ha conseguido crear un individuo —llamémoslo señor Marioneta— que, por como está diseñado, comete un delito, digamos que un asesinato cuando algo no marchó bien en un lío de drogas».[23] Y esto es lo que dicen sobre su experimento mental:

> Sí, es tan racional como otros criminales, y sí, fueron sus deseos y creencias los que produjeron sus acciones. Pero esos deseos y creencias estaban manipulados por fuerzas externas, y por eso, intuitivamente, merece nuestra compasión antes que nuestra condena moral […] ¿Cuál es la diferencia real entre nosotros y el señor Marioneta? Una evidente es que el señor Marioneta es víctima de una conspiración diabólica, mientras que la mayoría de la gente, suponemos, no. ¿Pero esto importa? El pensamiento de que el señor Marioneta no es completamente responsable depende de la idea de que sus acciones estuvieron determinadas externamente […] Pero el hecho de que estas fuerzas estén conectadas con los deseos e intenciones de los

malvados científicos es irrelevante, ¿o no? Lo que importa es únicamente que estas fuerzas están más allá del control del señor Marioneta, que realmente no son suyas.[24]

¿Qué te parece? ¿Es una buena bomba de intuición para lo que se propone o no? Como un detalle divertido, los autores observan que ésta es una cuestión que vale la pena considerar: «Daniel Dennett podría objetar que la historia del señor Marioneta no es más que una "bomba de intuición" engañosa». Ya lo creo que sí. Digo que es una muleta de razonamiento. Pero ellos no se dejan afectar por esta corazonada y siguen adelante: «Nos parece que mientras más sabe uno del señor Marioneta y de su vida, menos inclinado está a verlo como alguien verdaderamente responsable de sus acciones y a pensar que castigarlo sería un digno fin en sí mismo».

Entonces echemos un vistazo más de cerca, giremos las perillas de esta bomba de intuición y veamos qué es lo que realmente está haciendo el trabajo. Propongo que ajustemos cuatro perillas. Primero, eliminemos la conspiración diabólica, que los propios autores insisten en que no importa. Me sorprende que introduzcan tan despreocupadamente a un «neurocirujano nefando» y supongan que esto evidentemente es inocente, pues durante años se han expresado dudas sobre esa movida, pero podemos poner a prueba su convicción remplazando «un grupo de científicos» con «un medio ambiente indiferente»:

ANTES: Supongamos, entonces, que *un grupo de científicos* ha conseguido crear un individuo —llamémoslo señor Marioneta— que, por como está diseñado, comete un delito, digamos que un asesinato cuando algo no marchó bien en un lío de drogas.

DESPUÉS: Supongamos, entonces, que *un medio ambiente indiferente* ha conseguido crear un individuo —llamémoslo señor Marioneta— que, por como está diseñado, comete un delito, digamos que un asesinato cuando algo no marchó bien en un lío de drogas.

Segunda perilla: Si ya no tenemos a los conspiradores, tenemos que remplazar «por como está diseñado» con «con una alta probabilidad»:

ANTES: Supongamos, entonces, que un medio ambiente indiferente ha conseguido crear un individuo —llamémoslo señor Marioneta— que, *por como está diseñado,* comete un delito, digamos que un asesinato cuando algo no marchó bien en un lío de drogas.

DESPUÉS: Supongamos, entonces, que un medio ambiente indiferente ha conseguido crear un individuo —llamémoslo señor Marioneta— que, *con una alta probabilidad,* comete un delito, digamos que un asesinato cuando algo no marchó

bien en un lío de drogas.

Tercera perilla: quiero cambiar la motivación del crimen; sigue siendo un asesinato, pero en un escenario un poco distinto (eso no debería importar, ¿o sí?):

ANTES: Supongamos, entonces, que un medio ambiente indiferente ha conseguido crear un individuo —llamémoslo señor Marioneta— que, con una alta probabilidad, comete un delito, digamos que un asesinato *cuando algo no marchó bien en un lío de drogas*.

DESPUÉS: Supongamos, entonces, que un medio ambiente indiferente ha conseguido crear un individuo —llamémoslo señor Marioneta— que, con una alta probabilidad, comete un delito, digamos que un asesinato *para encubrir un desfalco*.

Cuarta perilla: cambiemos del nombre del acusado. Después de todo, no es más que un nombre:

ANTES: Supongamos, entonces, que un medio ambiente indiferente ha conseguido crear un individuo —llamémoslo *señor Marioneta*— que, con una alta probabilidad, comete un delito, digamos que un asesinato para encubrir un desfalco.

DESPUÉS: Supongamos, entonces, que un medio ambiente indiferente ha conseguido crear un individuo —llamémoslo *Capitán Autonomía*— que, con una alta probabilidad, comete un delito, digamos que un asesinato para encubrir un desfalco.

¿Ahora qué intuiciones te parece que bombea esto? ¿Las mismas? ¿Estás más inclinado a tenerle compasión que a condenarlo? A lo mejor ayudaría si les damos más cuerpo a los detalles. Aquí está mi amistosa enmienda a la bomba de intuición de Greene y Cohen:

El Capitán Autonomía se especializó en economía en Harvard y tras su graduación entró a trabajar en Lehman Brothers, donde toda la gente que lo rodeaba estafaba, y gracias a eso ganaba fuertes sumas de dinero. Se enamoró de una rompecorazones cazafortunas que amenazó con dejarlo si no se hacía rico pronto. Vio su oportunidad, un desfalco prácticamente indetectable, casi con seguridad un crimen invisible, y a sabiendas corrió el riesgo, pero, ¡ay!, aunque parezca increíble, surgió un testigo, que cometió el error de pararse demasiado cerca del barandal en la terraza del *penthouse*... Un rápido «tropezón» —oh oh—, y el testigo cayó muerto en la avenida. Se despertaron sospechas, y en poco tiempo el Capitán Autonomía fue arrestado.

¿Sigues inclinado a pensar que *porque sus acciones estaban determinadas externamente* él «no es verdaderamente responsable»? Incluso si estás tentado de ver al Capitán Autonomía como *víctima* de su ambiente (popof), creo que tendrás que aceptar que la tentación disminuye mucho, e incluso podría no ser más que una especie de vestigio de la narración anterior. (O quizá sólo he explotado tus emociones en la dirección opuesta, y cuento con tu sanguinario deseo de castigar a los avariciosos lobos de Wall Street por su papel en nuestro quebranto económico.) No digo que mis variaciones demuestren que la gente puede o no ser responsable a pesar de estar determinada; sólo digo que esta bomba de intuición específica no es nada confiable, pues los ajustes de las perillas (disponibles, permisibles) están interfiriendo demasiado con nuestros propios juicios. Puede no haberse diseñado para ocultar la verdad, pero ciertamente consigue dificultar la claridad de pensamiento.

Resumen

A la gente le preocupa profundamente tener libre albedrío, pero también parece tener ideas equivocadas sobre lo que el libre albedrío es o podría ser (como sus ideas equivocadas sobre el color y la conciencia). Nuestras decisiones no son pequeños milagros en el cerebro que transgreden la física y la química subyacente en el resto de nuestros procesos corporales, aunque muchos piensen que esto es lo que debe pasar si nuestras decisiones son verdaderamente libres. Sin embargo, de aquí no podemos concluir que *no* tenemos libre albedrío, porque el libre albedrío en este sentido descabellado no es el único concepto de libertad. La ley, de acuerdo con el sentido común, contrasta firmar un contrato «por tu propia voluntad» con firmarlo bajo coerción o bajo la influencia de alucinaciones o algún otro trastorno mental. Es éste un sentido de libre albedrío perfectamente familiar, una distinción que muchas de las prácticas y actitudes que conforman nuestra imagen manifiesta presuponen, y que no ha demostrado ninguna dependencia del sentido descabellado.

Durante cientos de años ha habido filósofos que insisten en que este sentido de libre albedrío es el importante, el que debería preocuparnos, y que es compatible con el determinismo, con el materialismo, con la física y la química que reinan sin que nadie las ponga en entredicho. Las bombas de intuición y las otras herramientas de pensamiento de esta parte están ideadas para apoyar y promover la comprensión de esta perspectiva, el *compatibilismo*. Ha tenido muchas versiones a lo largo de los años, y probablemente sea la opinión más generalizada no sólo entre filósofos sino también entre jueces, abogados y otros que tienen que hacer distinciones entre quién es responsable de qué y quién está eximido de culpa porque no tenía libre albedrío en el momento en que actuó. Algunos científicos están ahora poniendo en entredicho este consenso, y desde luego puede ser que hagan bien. Así, analicemos los argumentos seriamente.

A lo mejor la ciencia nos está enseñando algo radical, incluso revolucionario: que nadie es jamás responsable de nada de lo que hace, y que no hay una base razonable para distinguir algunos actos como dignos de elogio y otros como condenables. Para esa conclusión revolucionaria hace falta mucha más atención concienzuda a los detalles que la que hasta ahora ha recibido de los científicos que la sostienen. A la neurocirujana nefanda le bastó con una idea falsa para incapacitar a su paciente; una visión equivocada promulgada por científicos influyentes podría despojar a la gente de una variedad del libre albedrío legítima y que mejora la calidad de vida. Se requiere cautela de las dos partes.

El compatibilismo, con toda su popularidad entre los filósofos, siempre ha despertado sospechas. Es bien sabido que Immanuel Kant lo llamaba un «subterfugio mezquino», y hoy en día no faltan autores que expresan dudas sobre la sinceridad de quienes mantenemos esa postura. Así es como debería ser, de hecho. La ciencia nos enseña a

que estemos en guardia contra el pensamiento desiderativo, y muchas de las reglas de la investigación científica están especialmente ideadas para impedir que nuestras esperanzas nos engañen cuando creemos que son las pruebas lo que nos está convenciendo. Imagina que unos astrónomos anuncian que un asteroide gigante impactará con nuestro planeta dentro de 10 años y arrasará con toda forma de vida, y luego otro grupo de astrónomos sostiene que han vuelto a analizar los datos y podemos respirar tranquilos: el asteroide, por un estrecho margen, no chocará con la Tierra. Son buenas noticias, pero ¿cómo sabemos que los astrónomos del segundo grupo no se están engañando a sí mismos… o engañándonos a nosotros con una mentira piadosa? Revisa y vuelve a revisar sus cálculos; organiza un intento independiente de réplica. No nada más aceptes agradecido su conclusión porque no contiene ningún error evidente y te resulta atractiva. Pero tampoco olvides que es posible que tengan razón. No cometas el error opuesto de desacreditar —basado en «principios generales»— algo que parece «demasiado bueno para ser verdad».

¿Es el compatibilismo demasiado bueno para ser verdad? Creo que no. Creo que *es* verdadero y que podemos desestimar categóricamente a los alarmistas, y al mismo tiempo reformar y revisar nuestra comprensión de lo que avala nuestro concepto de responsabilidad moral. Pero ésa es una tarea para el futuro, y debería ser obra de muchas manos. Hasta donde puedo ver, es el problema filosófico más difícil y también el más importante de los que enfrentamos hoy en día. Hay mucho en juego, los temas son espinosos, y las emociones tienden a ofuscarnos. Vamos a necesitar todas nuestras herramientas de pensamiento, y otras que tendremos que fabricar sobre la marcha.

[1] D. Denys *et al.,* «Deep Brain Stimulation of the *Nucleus accumbens* for Treatment-Refractory ObsessiveCompulsive Disorder», *Archives of General Psychiatry,* 67 (10): 1061-1068, 2010.

[2] Si se duda que inculcar una creencia así pudiera tener tales consecuencias, véanse K. D. Vohs y J. W. Schooler, «The Value of Believing in Free Will: Encouraging a Belief in Determinism Increases Cheating», *Psychological Science,* 19 (1): 49-54, 2008, y la subsiguiente bibliografía para conocer las pruebas experimentales.

[3] T. Wolfe, «Sorry, but Your Soul Just Died», *op. cit.,* p. 100.

[4] W. Poundstone, *The Recursive Universe: Cosmic Complexity and the Limits of Scientific Knowledge,* William Morrow, Nueva York, 1985, p. 38.

[5] Para más sobre las implicaciones teóricas de esta compensación en espacio y tiempo, véase el cap. 9 de D. C. Dennett, *The Intentional Stance,* The MIT Press, Cambridge, 1987.

[6] Para una perspectiva completamente distinta de la física y la ingeniería en dos dimensiones, véase A. K. Dewdney, *The Planiverse: Computer Contact with a Two-Dimensional World,* Poseidon Press, Nueva York, 1984, una mejora sustancial con respecto a *Flatland* de E. A. Abbott, y una magnífica herramienta de pensamiento por derecho propio.

[7] W. Poundstone, *The Recursive Universe..., op. cit.,* pp. 227-228.

[8] J. von Neumann y O. Morgenstern, *Theory of Games and Economic Behavior,* Princeton University Press, Princeton, 1944.

[9] Un delicioso ejemplo de conjetura adelantada de nivel superior galopante aparece en la película *The Princess Bride,* en la escena en la que Wallace Shawn (Vizzini) trata de ser más listo que Cary Elwes (Westley) en el encuentro de «¿Cuál de las tazas tiene el veneno?» y termina siendo más listo que sí mismo. Véase <www. dailymotion.com/video/xhr71a_never-go-in-against-a-sicilian-when-death-is-on-the-line_shortfilms>.

[10] D. C. Dennett, *Breaking the Spell, op. cit.*

[11] L. Clegg, *Protean Free Will,* California Institute of Technology, Pasadena, 2012, ofrece un análisis formal pionero.

[12] J. Fodor, «Why Would Mother Nature Bother?», *The London Review of Books,* 25 (5): 17-18, 2003, p. 18.

[13] De acuerdo con el sitio web oficial de la Corona británica: «Hoy en día, la Corona conserva el derecho de propiedad de todos los cisnes sin marcar en aguas abiertas al público, pero la reina sólo ejerce su propiedad en ciertos trechos del Támesis y los afluentes que lo rodean. Esta propiedad se comparte con la Honorable Compañía de Viticultores y la Honorable Compañía de Tintoreros, a quienes la Corona les concedió derechos de propiedad en el siglo XV. Desde luego, hoy en día los cisnes ya no se comen».

[14] D. Wiggins, «Towards a Reasonable Libertarianism», en T. Honderich (comp.), *Essays on Freedom of Action,* Routledge & Kegan Paul, Londres, 1973, p. 54.

[15] G. Strawson, «Your Move: The Maze of Free Will», *The New York Times,* Opinionator, Nueva York, 22 de julio de 2010.

[16] *Idem.*

[17] G. Strawson, «Evolution Explains It All for You», *The New York Times,* Archives, Nueva York, 2 de marzo de 2003.

[18] D. Hofstadter, *Metamagical Themas..., op. cit.*

[19] D. Wooldridge, *The Machinery of the Brain,* McGraw-Hill, Nueva York, 1963, p. 82.

[20] S. Harris, *Free Will,* Free Press, Nueva York, 2012.

[21] H. Frankfurt, «Alternate Possibilities and Moral Responsibility», *Journal of Philosophy,* 65: 829-833, 1969.

[22] J. Greene y J. Cohen, «For the Law, Neuroscience Changes Everything and Nothing», *Philosophical*

Transactions of the Royal Society, 359: 1775-1785, 2004, p. 1776.

[23] *Ibid.,* p. 1780.

[24] *Idem.*

IX. ¿Cómo es ser un filósofo?

Cuando estamos demasiado cerca de algo, es difícil ver qué es. En los últimos años he pasado menos tiempo que antes interactuando con otros filósofos, y más tiempo interactuando con científicos y otras clases de pensadores. Sigo siendo un filósofo (¡no importa lo que algunos filósofos digan!), y me entusiasma la tarea de explicarles a los no filósofos por qué la filosofía es una actividad que vale la pena. Para esos que están completamente fuera de la filosofía, a menudo parece absurdo el paradigma mismo de la inteligencia inútil. Hay algo que no están viendo (recuerda la ley de Sturgeon: 90 por ciento de todo es basura; no están viendo el 10 por ciento). Al cabo de 50 años dentro de la filosofía, estoy familiarizado con ella; ahora que paso tanto tiempo lejos de la filosofía, también puedo ver vívidamente su lado extraño. Algunos de mis amigos y colegas científicos confiesan que de plano no entienden por qué no abandono el barco y me uno a ellos. La respuesta corta es que he conseguido, sentado a horcajadas en las fronteras, tener lo mejor de ambos mundos. Al trabajar con científicos obtengo una rica dieta de *datos* fascinantes y problemáticos sobre los cuales pensar, pero al seguir siendo un filósofo sin un laboratorio o una beca de investigación, consigo pensar sobre las teorías y sobre los experimentos sin tener que lavar nunca los platos. Para citar una de mis canciones favoritas de Gershwin, «es un lindo trabajo si puedes conseguirlo».

Me parece que en estos días, felizmente, los científicos están prestando una mayor y más respetuosa atención a los filósofos que antes, sobre todo en mi área principal, la mente, donde los fenómenos que estudian los cognitivistas son casi exactamente los mismos fenómenos en los que los filósofos han estado pensando durante siglos: la percepción, la memoria, el significado, la volición y la conciencia. Y los filósofos —algunos de ellos— se han ganado esa atención al informarse de la ciencia pertinente y concibiendo propuestas provechosas para aclarar y fomentar la investigación científica, y al encontrar mejores maneras de explicar los resultados al mundo no científico. No obstante, sigue habiendo importantes fallos de comunicación cuando las tribus tratan de interactuar; a continuación abordaré algunas de las diferencias con el propósito de conseguir un mayor entendimiento mutuo en el futuro.

74. Un trato faustiano

Durante varios años he estado planteándoles la siguiente pregunta a mis colegas filósofos: si Mefistófeles te ofreciera las siguientes dos opciones, ¿cuál elegirías?

> *A*. *Resuelves* el gran problema filosófico de tu elección de manera tan concluyente que ya no hay nada más que decir (gracias a ti, parte del campo cierra para siempre, y obtienes un pie de página en la historia).
>
> *B*. Escribes un libro que seduce por su perplejidad y controversia, que permanece en la lista de lecturas obligatorias en los siglos venideros.

Algunos filósofos reconocen a regañadientes que se inclinarían por la opción *B*. Si tuvieran que elegir, preferirían tener lectores que tener razón. Como los compositores, los poetas, los novelistas y otros creadores artísticos, tienden a querer que millones (¡miles de millones, si es posible!) de personas experimenten su obra, una y otra vez. Pero también se sienten atraídos hacia la búsqueda del científico. Después de todo, se *supone* que los filósofos tratan de llegar a la verdad.

Cuando les presento el mismo trato faustiano a los científicos, tienden a optar por *A* sin vacilación: para ellos es algo obvio. Y luego mueven la cabeza en señal de asombro (¿o disgusto?) cuando se enteran de que para muchos filósofos es una decisión difícil y de que muchos optan, un poco tímidamente, por *B*. Pero esta reacción de los científicos pierde de vista la importante observación que hizo Nicholas Humphrey (véase el apartado 48):

> En *Las dos culturas,* C. P. Snow ensalzó los grandes descubrimientos científicos como «Shakespeare de la ciencia», pero en un sentido estaba fundamentalmente equivocado. Las obras de Shakespeare eran de Shakespeare y de nadie más; por el contrario, los descubrimientos científicos no pertenecen —en última instancia— a nadie en particular.[1]

Si Shakespeare no hubiera existido, nadie más habría escrito *Hamlet, Romeo y Julieta* o *El rey Lear*. Si Van Gogh no hubiera existido, nadie más habría pintado la *Noche estrellada*. Puede ser una leve exageración, pero tiene algo de cierto. Por un lado, en las aportaciones de los grandes artistas hay una individualidad que no sólo parece ser excepcional en la ciencia, sino decididamente fuera de lugar. Las famosas polémicas científicas sobre quién fue primero, y las carreras por un factor decisivo para ganar un premio Nobel u otro, son feroces precisamente porque alguien más *podría* haber hecho

exactamente la misma aportación que tú te estabas esforzando en hacer —y no te van a conceder puntos por tu estilo si llegas en segundo lugar—. Estas competencias no tienen paralelo en las artes, donde reina un diferente conjunto de objetivos.

Algunos científicos aspiran a alcanzar muchos lectores, y a deleitar a los que atrapen, y los mejores escriben obras de un valor literario incomparable. Vienen a la mente los libros de Darwin. Pero la finalidad de estar en lo correcto, de convencer a los lectores de una verdad descubierta, sigue siendo lo más importante, como puede apreciarse de un vistazo si comparamos el *Diario de viaje de un naturalista alrededor del mundo en el S. M. Beagle* de Darwin con *Moby Dick* de Melville. Puede aprenderse mucho sobre ballenas y su caza leyendo *Moby Dick,* pero Melville no la escribió para que fuera un ingenioso y persuasivo compendio de los hechos sobre la caza de ballenas. Teniendo presente la diferencia entre las finalidades de la ciencia y las finalidades del arte, pues, he aquí una pregunta para científicos que es un paralelo adecuado de la pregunta provocadora que les hago a mis colegas filósofos. Si Mefistófeles te presentara las siguientes dos opciones, ¿cuál elegirías?:

1. Ganas la carrera (y el acompañante premio Nobel) por haber hecho un descubrimiento que se convierte en la base para una enorme expansión del conocimiento científico pero que, en retrospectiva, era la personificación de la sentencia de Humphrey, que no pertenece a nadie en particular. (Vienen a la mente Crick y Watson; hay pocas dudas de que si no hubieran ganado la carrera cuando lo hicieron, Linus Pauling o alguien más lo habría hecho pronto.)

2. Propones una teoría tan original, y tan inimaginable antes de tu trabajo, que tu apellido se incorpora a la lengua, pero tu teoría resulta estar en su mayor parte equivocada, aunque por años —siglos, incluso— ha seguido generando una valiosa polémica. (Pienso en el dualismo cartesiano mente-cuerpo, las teorías de la evolución de Lamarck, el conductismo de Skinner y las posturas freudianas sobre todas las cosas, desde la sexualidad infantil y la neurosis hasta el arte, la música y la literatura.)

Un mejor ejemplo, aunque menos conocido, de la opción *2* podría ser el ambicioso trabajo de Descartes sobre la física, que fue tan influyente y estuvo tan brillantemente equivocado que resultó una gran provocación para Isaac Newton, cuyo *Philosophiae Naturalis Principia Mathematica* (escrito en 1687), que cambió el mundo, deliberadamente le hacía eco al título de los *Principia Philosophiae* de Descartes (de 1644) para dejar claro qué cosmovisión pretendía remplazar. Y luego está la lingüística chomskiana. Sin duda pasa la prueba de originalidad. Como la victoria del *America* en la carrera que le dio nombre a la competencia de veleros America's Cup, no se veía un

segundo lugar por ningún lado cuando Chomsky irrumpió en la escena. En los años subsecuentes, la semilla teórica original —la teoría de la «gramática generativa transformacional» de las estructuras sintácticas (en *Syntactic Structures,* publicado en 1957)— se ha abandonado casi por completo, y la han remplazado teorías descendientes de varias especies tan diferentes de su ancestro común como las avestruces, los colibríes y los albatros de los dinosaurios a partir de los cuales evolucionaron. ¿Estaba Chomsky fructíferamente equivocado en 1957, o **más bien** (¡ring!) descubrió una gran verdad? «Sí» responde muy bien la pregunta.

Honramos a los científicos que se equivocan de maneras provechosas (recuerda el insulto de Wolfgang Pauli sobre el teórico que «ni siquiera estaba equivocado»). Sin embargo, si te obligaran a elegir, ¿cambiarías ser el primero y tener razón por ser original y provocador? No es tan fácil decidirse, ¿verdad?

75. La filosofía como una autoantropología naïf

Patrick Hayes, el investigador de inteligencia artificial, una vez emprendió un proyecto para axiomatizar la física naïf (o vulgar) de líquidos. La idea era proporcionarle a un robot las proposiciones que habría de usar como creencias básicas si interactuaba con personas (que utilizan la física vulgar todos los días). Demostró ser más difícil de lo que había previsto, y escribió un interesante artículo sobre el proyecto, «The Naïve Physics Manifesto».[2] En la física naïf de líquidos, todo lo que a las personas naïf o ingenuas les parece antiintuitivo está, desde luego, descartado: los sifones son «imposibles» y también las pipetas, pero puedes secar líquidos con una toalla esponjosa y sacar agua del pozo con una bomba de succión. Un robot equipado con esa reserva de «conocimiento» se sorprendería tanto con un sifón como la mayoría de nosotros la primera vez que vimos uno en acción. El proyecto de Hayes era lo que yo llamaría física naïf *sofisticada,* porque no se hacía ilusiones: sabía que la teoría que intentaba axiomatizar era falsa, por útil que fuera en la vida diaria. Era un ejercicio en lo que podría llamarse antropología axiomática: trata lo que dice la gente —y con lo que estás tú de acuerdo— como tus axiomas o teoremas, y trata de hacer que los datos sean consistentes, resolviendo cualquier contradicción que encuentres. Y, desde luego, no se molestaba en congregar informantes reales; supuso que sabía de física naïf de líquidos tanto como cualquier persona normal, así que se usó a sí mismo como su único informante: *auto*antropología axiomática.[3]

Ahora compara el proyecto de Hayes con los proyectos filosóficos en *metafísica analítica,* que a veces me dan la impresión de ser una autoantropología naïf *naïf*, pues los participantes en este tipo de investigación parecen convencidos de que su programa realmente llega a algo verdadero, no sólo a algo que cree verdadero determinada subclase de seres humanos (los filósofos anglófonos convencidos de la metafísica analítica). Por lo demás, los programas parecen idénticos: juntas tus intuiciones compartidas, las pruebas y las provocas entrando en un mutuo bombeo de intuiciones, y luego maquillas el conjunto de datos resultantes para que parezca una «teoría» consistente, basada en principios «aceptados» que cuentan, idealmente, como axiomas. Les he preguntado a algunos metafísicos analíticos si pueden distinguir entre su empresa y la autoantropología naïf naïf de su clan, y hasta ahora no he recibido ninguna respuesta convincente.

La alternativa es la antropología naïf *sofisticada* (tanto la *auto* como la *hetero*): la antropología que se reserva la opinión sobre si cualquiera de los teoremas aducidos merece que se confíe en él. Éste es un proyecto viable y a menudo valioso. Propongo que ésta es la empresa a la que deberían dedicarse los metafísicos analíticos, pues requiere unos ajustes más bien mínimos a sus métodos y una sola revisión a fondo de su razón de ser: deben bajarles a sus pretensiones y reconocer que su investigación se ve

mejor como un reconocimiento preparatorio del terreno de la imagen manifiesta, que suspende tanto la creencia como la incredulidad, como hacen los antropólogos cuando estudian una cultura exótica: pretendamos por el momento que los nativos tienen razón, y veamos qué resulta. Como al menos una gran parte de la tarea de la filosofía, en mi visión de la disciplina, consiste en gestionar el tráfico de un lado a otro entre la **imagen manifiesta** y la **imagen científica**, es buena idea que los filósofos analicen con qué suposiciones vulgares están compitiendo antes de ponerse a elaborar unas teorías y a criticar otras.

Uno de los sellos distintivos de la antropología naïf sofisticada es su apertura a los descubrimientos antiintuitivos. Siempre que estés haciendo antropología naïf, la antiintuición (hacia los nativos) perjudica tu reconstrucción; cuando cambias de equipamiento y empiezas a preguntar qué aspectos de la «teoría» naïf son verdaderos, lo antiintuitivo pierde fuerza como objeción e incluso, en ocasiones, se vuelve una señal de progreso significativo. En la ciencia en general, los resultados antiintuitivos son muy valorados, después de todo.

Una de las debilidades de la autoantropología es que nuestras propias intuiciones se prestan a que nuestras predilecciones teóricas las distorsionen. Los lingüistas han sabido por mucho tiempo que llegan a estar tan metidos en sus teorías que dejan de ser fuentes confiables de intuición lingüística. ¿De verdad puedes decir en español «El niño el hombre la mujer besó golpeó huyó», o mis estudios sobre los constituyentes sintácticos están engañando a mi «oído»? Sus intuiciones crudas, indoctas, se han mancillado con tanta teoría, así que reconocen que deben salir y preguntarles a los no lingüistas sobre sus intuiciones lingüísticas. Recientemente los filósofos han empezado a darse cuenta de esto, en el nuevo entusiasmo por la llamada filosofía experimental.[4] Aún es pronto para saberlo, y algunos de los esfuerzos pioneros no son admirables, pero al menos los filósofos están haciéndose a la idea de que ya no pueden declarar que diversas proposiciones son evidentemente verdaderas bajo el único argumento de que les parecen fantásticamente evidentes *a ellos*. (En una vena similar, Hayes *podría* haberse sorprendido con los principios centrales de la física vulgar si se hubiera tomado la molestia de entrevistar a una muestra aleatoria de gente en lugar de solamente tratarse a sí mismo como modelo.)

Entonces aquí hay un proyecto, un tipo particular de antropología naïf sofisticada, que los filósofos deberían considerar seriamente para emprender una investigación del terreno de la imagen manifiesta o de sentido común del mundo antes de ponerse a elaborar sus teorías del conocimiento, la justicia, la belleza, la verdad, la bondad, el tiempo, la causalidad, etc., para asegurarse de que realmente dirigen sus análisis y argumentos a modelos que le importan al resto del mundo, tanto preocupaciones legas como preocupaciones científicas. Esta investigación sistemática daría lugar a algo

parecido a un catálogo del terreno conceptual no reformado que le señala los problemas al teórico, la metafísica de la imagen manifiesta, si tú quieres. Es aquí donde los filósofos tenemos que empezar en nuestros intentos de gestionar de un lado a otro entre las últimas innovaciones en la imagen científica, y no dolería tener un mapa cuidadoso de este territorio vulgar en lugar de sólo mirarlo de arriba abajo. Ésta es la otra mitad, podría decirse, de la reforma que transformó la filosofía de la ciencia, de un campo de fantasía de café, a una seria asociación con la ciencia real, cuando los filósofos de la ciencia decidieron que en realidad tenían que saber mucho de ciencia actual desde dentro. Una vez que pensemos sobre nuestras tareas filósoficas con esta imagen en mente, veremos que gran parte de esas acciones informales, consistentes en avanzar dificultosamente, mostrarse indeciso, sembrar contraejemplos y degradar intuiciones, con que se llenan las páginas de las revistas de filosofía, es, en el mejor de los casos, un intento de organizar un consenso mutuamente aceptable sobre este territorio.

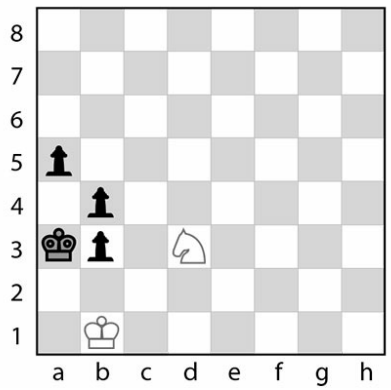
Analiza este acertijo de ajedrez.[5] Blanca a jaque mate en dos (figura 24).

Apareció recientemente en el *Boston Globe,* y lo que me asustó fue que yo creía que ya se había demostrado que no puedes hacer jaque mate con un caballo solitario (y un rey, desde luego). Me equivocaba; como David Misialowski me señaló recientemente en un correo, se ha demostrado que no puedes darle mate al rey de tu oponente cuando sólo su rey, tu rey y tu caballo están en el tablero. El hecho de que *no* sea una verdad del ajedrez la proposición *Nunca puedes hacer jaque mate con un caballo solitario y el rey,* es una verdad del ajedrez de orden superior.

La filosofía es una disciplina tradicionalmente *a priori,* como las matemáticas, o al menos tiene en su núcleo una metodología *a priori,* y este hecho es una navaja de dos

FIGURA 24

BLANCAS JUEGAN
Pista: jaque mate en 2 jugadas.



Solución: 1. Cb2 a4 (la única jugada legal) 2. Cc4 mate (de un estudio de Gurevich).

filos. Por un lado, libra a los filósofos de pasar horas tediosas en el laboratorio o afuera en el campo, y de tener que aprender técnicas de recolección de datos, métodos estadísticos, geografía, historia, lenguas extranjeras, ciencia empírica, etc., así que tienen mucho tiempo para afilar sus habilidades filosóficas. Por otro lado, como muchas veces se ha observado, la filosofía puede crearse prácticamente de cualquier cosa, y esto no siempre es bueno. Para los lectores más jóvenes que estén pensando en dedicarse profesionalmente a la filosofía, este capítulo quiere advertirles que la libertad y el carácter abstracto propios de esta disciplina pueden ser desventajas. Este capítulo es también una guía de viaje para mostrarles a los fuereños algo de las tradiciones y de los escollos de la filosofía.

Piensa en las verdades del ajedrez como paradigma de las verdades *a priori*. Es un hecho empírico que la gente juega al ajedrez, y hay montañas de otros datos empíricos acerca de él; por ejemplo, sobre cómo la gente lo ha estado jugando durante siglos, cómo a menudo usa piezas magníficamente talladas en tableros con incrustaciones, etc. El conocimiento de ninguno de estos hechos empíricos desempeña ningún papel indispensable en la actividad de encontrar las verdades *a priori* del ajedrez, que también

las hay en abundancia. Todo lo que tienes que saber son las reglas del juego. Hay exactamente 20 jugadas de apertura legales (16 de peón y cuatro de caballo); un rey y un alfil solitario no pueden dar jaque mate a un rey solitario, y tampoco un rey y un caballo solitario, etc. Encontrar estas verdades *a priori* del ajedrez no siempre es fácil. Tan sólo demostrar qué es y qué no es posible dentro de las reglas del ajedrez es una tarea complicada, y se pueden cometer errores. Por ejemplo, hace pocos años un programa de ajedrez de computadora descubrió una red de mate —una victoria garantizada o forzada — consistente en más de 200 jugadas sin una captura. Esto desmintió un «teorema» del ajedrez que estaba vigente desde hacía tiempo y ha obligado a un cambio en las reglas del juego. Decía que 50 jugadas sin una captura por cualquiera de las partes constituía un empate (acabar en tablas), pero como esta prolongada red de mate es irrompible, y conduce a una victoria, no es razonable mantener las tablas de 50 jugadas. (Antes de que las computadoras empezaran a jugar al ajedrez, nadie imaginaba que *pudiera* haber una victoria garantizada de esta duración ni mucho menos.) Todo esto puede ser muy interesante, y muchas personas de gran inteligencia han consagrado sus mentes a investigar este sistema de las verdades *a priori* del ajedrez.[6]

Algunos proyectos de investigación filosófica —o problemáticas, para decirlo más elegantemente— son un poco como entender las verdades del ajedrez. Se presupone un conjunto de reglas sobre las que hay un mutuo acuerdo —y que rara vez se discuten—, y a continuación se buscan, se articulan, se debaten, se refinan las implicaciones de esas reglas. Hasta aquí todo muy bien. El ajedrez es un artefacto humano profundo e importante, sobre el que se han escrito muchas cosas valiosas. Sin embargo, algunos proyectos de investigación filosófica se parecen más a tratar de encontrar las verdades del *ajetrez.* El ajetrez es como el ajedrez excepto porque aquí el rey puede avanzar dos escaques en cualquier dirección, no uno. Acabo de inventármelo, aunque seguramente otros lo han explorado a fondo a ver si vale la pena jugarlo. Probablemente no. Probablemente tenga otros nombres. No me tomé la molestia de investigar estas preguntas porque, aunque tienen respuestas verdaderas, simplemente no vale la pena dedicar tiempo y energía a averiguarlo. Al menos eso pienso. Hay tantas verdades *a priori* del ajetrez como del ajedrez (una infinidad), y son igualmente difíciles de descubrir. Y eso significa que si la gente se pusiera a investigar en serio las verdades del ajetrez, cometería errores, que necesitarían corregirse, y esto abre un campo completamente nuevo de investigación *a priori,* las verdades de *orden superior* del ajetrez, como las siguientes:

1. La prueba de Jones (1989) de que *p* es una verdad del ajetrez presenta fallos: él pasa por alto la siguiente posibilidad…
2. La afirmación de Smith (2002) de que la prueba de Jones (1989) presenta fallos presupone la verdad del lema de Brown (1975), que recientemente ha

puesto en duda Garfinkle (2002)…

Ahora bien, nada de esto es juego de niños. De hecho, a lo mejor uno podría manifestar una brillantez considerable en la actividad grupal de entender las verdades de orden superior del ajetrez. Aquí es donde viene muy a cuento la sentencia del psicólogo Donald Hebb: «Si no vale la pena hacerlo, no vale la pena hacerlo bien». Probablemente a todos los filósofos les vendrá rápido a la mente alguna polémica vigente en filosofía cuyos participantes se quedarían sin trabajo si la sentencia de Hebb se aplicara implacablemente, pero es seguro que no todos estaremos de acuerdo en precisamente cuáles industrias artesanales deberían cerrarse. Probablemente no haya una sola investigación en nuestra espaciosa disciplina de la que alguna escuela de pensamiento no sostenga que es un esfuerzo inútil, inteligencia derrochada en sacar a relucir los trapitos sucios de los demás. Una votación no arrojaría resultados a los que valiera la pena prestar atención, y decidirlo dictatorialmente sería todavía peor, así que deja que 1 000 flores se abran, digo yo. Sólo recuerda: si dejas que 1 000 flores se abran, cuenta con que 995 se marchiten. La advertencia que quiero darte es ésta: trata de evitar comprometer tus valiosos años formativos en un asunto de investigación de poca durabilidad. Las modas filosóficas pasan rápidamente, y algo de verdad debe haber en la regla general que dice: mientras más candente el tema, más pronto se quema.

Una buena prueba para asegurarse de que un proyecto filosófico no está más que explorando las verdades de orden superior del ajetrez es ver si la gente, fuera de los filósofos, realmente puede jugar la partida. ¿Puede conseguirse que a alguien ajeno a la filosofía académica le *interese* si el contraejemplo de Jones funciona contra el principio de Smith? Otra prueba es tratar de enseñarlo a estudiantes de licenciatura no iniciados. Si no «lo entienden», realmente deberías considerar la hipótesis de que estás siguiendo a un grupo de expertos autosuficientes hacia una trampa artificial.

Una manera como funciona la trampa es ésta: la filosofía, hasta cierto punto, no es un acto natural, y mientras más inteligente seas, más dudas y reservas es probable que tengas sobre si estás entendiéndola, si lo «estás haciendo bien», si tienes algún talento para la disciplina, e incluso si vale la pena entrar a la disciplina, para empezar. Así, *como corresponde,* el estudiante brillante Jones no está seguro de si debe hacer un posgrado en filosofía. Intrigado por la discusión del profesor Brown, Jones hace un intento, escribe un ensayo sobre el tema candente *H* y el profesor Brown le pone un 10. «Tienes mucho talento, Jones», le dice Brown, y Jones acaba de descubrir algo que podría ser un trabajo adecuado para toda la vida. Jones empieza a dedicar tiempo a aprenderse las reglas de este juego particular y a jugarlo ferozmente con los otros jóvenes aspirantes. «Vaya, ¡somos buenos para esto!», dicen, incitándose unos a otros. Si hay dudas sobre ciertas afirmaciones que hacen posible el proyecto, éstas tienden a acallarse o sofocarse «por

mor del argumento». A continuación vienen las publicaciones.

Así que no cuentes con la validación de tus compañeros en el posgrado *o* de tu profesor favorito para resolver la cuestión. Todos tienen un interés personal en que se mantenga en marcha el proyecto. Es lo que saben hacer; es para lo que son buenos. También en otros campos se presenta este problema, y posiblemente sea incluso más difícil escaparse de ahí. Los investigadores en alguna ciencia experimental que para ejercerla tienen que dominar una técnica y equipar un costoso laboratorio, a veces se quedan atorados llenando los espacios en blanco de matrices de datos que ya a nadie le importan. ¿Qué se pretende que hagan? ¿Tirar todos esos aparatos caros? Puede ser un problema desagradable. Para los filósofos realmente es más fácil y más barato reequiparse. A fin de cuentas, nuestro «entrenamiento» no es, por lo general, de alta tecnología. Se trata sobre todo de aprender a orientarse en diversas bibliografías, aprender las jugadas que ya se han intentado y puesto a prueba. Y aquí la trampa que hay que evitar es ésta: ves que alguien eminente ha publicado una afirmación insostenible o discutible; el artículo inteligente pero fallido del profesor Regadas es un blanco facilísimo, el objetivo ideal de un artículo vistoso para hacer tu debut. ¡Adelante! Intervienes en la discusión, junto con otras 12 personas, y ahora tienes que ver por dónde caminas, porque para cuando todos se hayan citado entre sí y hayan respondido a las respuestas, serás un experto en ciernes sobre cómo tratar la manera de tratar las respuestas a la menor exageración de Regadas. (Y recuerda también que si Regadas no hubiera hecho su tesis un poquito atrevida, nunca habría atraído la atención de tanta gente, para empezar; la tentación de ser provocador no es exclusiva de los estudiantes de posgrado en busca de una ostentosa entrada a la cancha.)

A algunas personas les satisface encontrar un grupo agradable de gente lista con la cual compartir «la diversión del descubrimiento, los placeres de la cooperación y la satisfacción de llegar a acuerdos», como lo expresó el filósofo John Austin en una conferencia que después fue publicada, y ni se preocupan sobre si vale o no la pena la tarea conjunta.[7] Y si suficientes personas se ponen a la tarea, tarde o temprano se convierte en un fenómeno por derecho propio que vale la pena estudiar. Como el filósofo Burton Dreben les decía a sus alumnos de posgrado en Harvard: «La filosofía es basura, pero la historia de la basura es erudición». Sin embargo, algunas basuras son más importantes que otras, y es difícil decidir cuáles puedan ser merecedoras de becas. En otra conferencia que se publicó en el mismo libro, Austin nos dio la siguiente obra maestra del sarcasmo: «No es raro que entre el público de una conferencia haya algunos que prefieran que las cosas sean importantes, y a ellos, si acaso hay alguno aquí presente, les debo una perorata».[8] Austin era un filósofo brillante, pero la mayoría de los muy prometedores filósofos que orbitaban a su alrededor, y que seguramente se rieron con este comentario, se han desvanecido sin dejar rastro: su inteligentísimo trabajo sobre

la filosofía del lenguaje ordinario (una escuela que Austin más o menos inventó) fue publicado, como cabía esperar, y luego, pocos años después de su publicación, fue completa y merecidamente olvidado. Ha pasado muchas veces.

Entonces ¿qué debes hacer? Las pruebas que he mencionado —ver si se consigue interesar en el tema a gente ajena a la filosofía o alumnos de licenciatura brillantes— sólo dan señales de alerta; no son definitivas. Seguramente ha habido, y habrá, temas prohibitivamente abstrusos y difíciles de investigación filosófica a los que bien valdría la pena dedicarse, por mucho que a un iniciado no lo impresionen. Para nada quiero desanimar exploraciones que contravengan las suposiciones ambientales sobre lo que es interesante e importante. Por el contrario, los mejores atrevimientos del campo casi siempre se toparán con fría incredulidad o burla al principio, y eso no debería desalentarte. Lo que digo es que no deberías subirte al tren e instalarte ahí con suficiencia sólo porque has encontrado a algunos otros pasajeros brillantes a los que les parece que tu trabajo en el tema es tan imposible de ignorar como a ti te parece el suyo. Puede ser que se estén tomando el pelo unos a otros.

77. El diez por ciento que está bien

Así, si la **ley de Sturgeon** se aplica a la filosofía como a todo lo demás, ¿cuál es, en mi opinión, la parte buena? Para empezar, los clásicos realmente lo son por alguna razón. De Platón a Russell, las lecturas habituales en los cursos de historia de la filosofía se mantienen firmes después de siglos de estudiarlas, y lo mejor de la bibliografía secundaria sobre esta bibliografía primaria también es muy valioso. Obtendrás *algo* —en realidad mucho— de leer a Aristóteles, a Kant o a Nietzsche por tu cuenta, sin lecturas complementarias, pero obtendrás mucho más si aceptas alguna orientación de quienes se han especializado en estos pensadores a lo largo de toda su carrera.

No todos los historiadores de la filosofía tienen las mismas metas y actitudes, y yo, sin ir más lejos, no veo ninguna buena razón para descalificar a ninguno de los contendientes. Algunos insisten en colocar a sus pensadores en el contexto histórico en el que escribieron, lo que significa, por ejemplo, aprender mucho sobre ciencia del siglo XVII si realmente quieres entender a Descartes, y mucho de historia política de los siglos XVII y XVIII si realmente quieres entender a Locke o a Hume, y siempre, desde luego, también mucha de la filosofía de sus contemporáneos menores. ¿Por qué molestarse con los del montón? Hay una buena razón. Me di cuenta de que nunca aprecié realmente a muchos de los pintores de los siglos XVI y XVII hasta que visité museos europeos donde pude ver sala tras sala llenas de pinturas de segunda categoría de los mismos géneros. Si lo único que ves es lo bueno —que es todo lo que ves en los cursos introductorios de sondeo, y en los mejores museos—, es muy difícil ver cuán maravilloso es. ¿Sabes cuál es la diferencia entre una buena biblioteca y una gran biblioteca? Una buena biblioteca tiene todos los buenos libros. Una gran biblioteca tiene todos los libros. Si realmente quieres entender a un gran filósofo, tienes que pasar algún tiempo mirando a sus contemporáneos y predecesores menores que quedaron a la sombra de los maestros.

Otros especialistas apenas si tocan los contextos históricos en los que trabajaron sus héroes, y mejor se concentran en mostrar cómo traducir las ideas a los contextos de hoy. Después de todo, Leibniz no escribió la *Monadología* para que fuera una obra ejemplar del racionalismo del siglo XVII; la escribió para *llegar a la verdad.* Al final, no te tomarás en serio a ningún filósofo hasta que preguntes si lo que dijo es o no *correcto.* Los estudiantes —y los profesores— de filosofía a veces lo olvidamos: nos concentramos en clasificar y nos ocupamos de «comparar y contrastar», como decimos en las preguntas de examen. Departamentos de filosofía enteros a veces caen en esta visión de su objetivo. Eso no es filosofía: sólo es *apreciación de la filosofía.* Así es como trato de ayudar a mis alumnos a romper este hábito:

> Has tropezado con un secreto terrible: una conspiración para destruir la Estatua de la Libertad, digamos, o para colapsar la red eléctrica nacional. Trabajas febrilmente para

recopilar y ordenar tus pruebas, y luego escribes una carta en la que haces gala de toda tu elocuencia. Mandas copias a la policía, el FBI, el *New York Times,* CNN, y recibes esta respuesta: «Ah, otro ejemplo muy inteligente de teoría de la conspiración post 11 de septiembre», y «Una lectura apasionante, de hecho, y bastante verosímil, a su manera, con excelentes trazos», y «Me recuerda a Dom de Lillo, con ecos de Pynchon». ¡Aaaarg! ¡Presten atención! ¡Estoy tratando de decirles la *verdad*! Respeta al filósofo al que estás leyendo preguntándote a ti mismo, con respecto a toda oración y párrafo, «¿Creo esto? Y si no, ¿por qué no?»

Además de la historia de la filosofía, hay excelentes trabajos en filosofía de la ciencia: matemáticas, lógica, física, biología, psicología, economía, teoría política. Casi no hay obras sobre la filosofía de la química, la astronomía, la geología o la ingeniería como tales, pero hay trabajos buenos sobre algunos de los temas conceptuales que surgen en estos campos. Y luego tenemos la ética. En 1971 John Rawls publicó *Teoría de la justicia,* una obra imponente que inauguró una era fructífera de filósofos que abordan los temas tradicionales de la ética con un ojo puesto en las ciencias sociales, especialmente la economía y la ciencia política, pero también en la biología y la psicología. En gran parte gracias a Rawls, los filósofos que trabajan en la ética mejoraron su juego, y el resultado ha sido una gran cantidad de investigación filosófica valiosa, que merece, y recibe, atención de los investigadores de otras disciplinas, y de los políticos y los críticos sociales.

Finalmente, hay filósofos que no son en absoluto interdisciplinarios, y también se apoyan ligeramente en la historia de su campo; se especializan en problemas contemporáneos que surgen en la obra de los filósofos que se especializan en los problemas contemporáneos que surgen en la obra de otros filósofos contemporáneos. Algo de esto, como ya he señalado, sucumbe a la regla de Hebb: si no vale la pena hacerlo, no vale la pena hacerlo bien. Pero una parte es excelente y valiosa. En este libro he mencionado a bastantes filósofos contemporáneos, y no lo haría si no pensara que sus ideas merecen ser consideradas seriamente, sobre todo cuando sostengo que están cometiendo un error. Aparte de mis objetivos, hay varias decenas de otros filósofos cuya obra admiro particularmente, pero no voy a cometer el error de anotarlos en una relación. Varias veces en mi carrera me he fiado del juicio de algún colega que me dijo que no me molestara con la obra de *X* porque es basura, sólo para enterarme tiempo después de que me dejé engañar e ignoré a un pensador con ideas valiosas cuya aportación a mi propio pensamiento se retrasó por culpa del mal dato. Tengo plena conciencia de cuán fácil es interpretar que estoy eximiendo a pensadores interesados en la lectura de algún filósofo que no llegó a mi lista. Así, pues, por favor trata este libro como una introducción con final abierto a *algunas* maneras de hacer filosofía, y si te

parecen provechosas, pueden ser un trampolín hacia tu propia exploración de las preguntas y respuestas en las que tantos pensadores han estado trabajando durante tanto tiempo.

[1] N. Humphrey, «Scientific Shakespeare», *op. cit.*

[2] P. Hayes, «The Naïve Physics Manifesto», en D. Michie (comp.), *Expert Systems in the Microelectronic Age,* Edinburgh University Press, Edimburgo, 1978.

[3] La obra de Hayes inspiró a otros. Véase la antología compilada por D. Bobrow, *Qualitative Reasoning about Physical Systems,* The MIT Press, Cambridge, 1985, p. 13. Los autores de uno de los ensayos allí incluidos comentan: «*Física naïf* es en sí mismo un término ambiguo. ¿Es simplemente mala física? ¿Es psicología? ¿Inteligencia artificial? ¿Física?». La respuesta, sostengo, es que la física naïf es el intento de hacer que la parte de la física que pertenece a nuestra imagen manifiesta sea lo suficientemente rigurosa para apoyar el razonamiento deductivo automatizado.

[4] Véase J. Knobe y S. Nichols (comps.), *Experimental Philosophy,* Oxford University Press, Oxford, 2008.

[5] Con respecto a otro acertijo en este libro, en primer lugar, «the *slop* I *love* is *even* followed by Laura *Dern*» [«a la bazofia que amo hasta la sigue Laura Dern»], y en segundo lugar, «in spite of the *smut,* I *hope* to *iron* Sean *Penn*'s shirt» [«a pesar de la inmundicia, espero planchar la camisa de Sean Penn»].

[6] El día de los inocentes de 2012 apareció una deliciosa broma de ajedrez («Rajlich: Busting the King's Gambit», *ChessBase New,* 2 de abril de 2012, <http://chessbase.com/newsdetail.asp?newsid=8047>. Decía que un consorcio gigante de computadoras que había trabajado día y noche a lo largo de cuatro meses finalmente había demostrado que hay una condición, si bien poco probable, en la que es imposible derrotar el gambito de rey (una apertura que por mucho tiempo ha estado desprestigiada pero nunca se ha «refutado»). Me dejé engañar, cosa especialmente vergonzosa pues antes había escrito sobre una broma de Martin Gardner en la introducción de M. Hurley, D. C. Dennett y R. B. Adams Jr. a *Inside Jokes: Using Humor to Reverse-Engineer the Mind,* The MIT Press, Cambridge, 2011. Para la explicación de cómo se creó la broma, véase «The ChessBase April Fools Prank», *ChessBase News,* 4 de abril de 2012, <http://chessbase.com/newsdetail.asp?newsid=8051>.

[7] J. L. Austin, *Philosophical Papers,* Oxford University Press, Oxford, 1961, p. 123.

[8] *Ibid.,* p. 179.

X. Usa las herramientas. Esfuérzate más

«¡Es inconcebible!» Eso es lo que algunas personas declaran cuando se enfrentan al «misterio» de la conciencia, o a la afirmación de que la vida surgió en este planeta hace más de 3 000 millones de años sin ayuda de ningún Diseñador Inteligente, por ejemplo. Cuando oigo la exclamación, siempre me siento tentado de decir: «Bueno, desde luego que *para ti* es inconcebible. Dejaste olvidadas tus herramientas de pensamiento y ni siquiera estás haciendo un esfuerzo». Recuerda la firme declaración de William Bateson de que una base material para los genes era inconcebible. Hoy en día ni siquiera a los niños de primaria les cuesta trabajo concebir el ADN, y no porque sean más brillantes que Bateson: es porque en el último siglo hemos ideado y pulido las herramientas de pensamiento que facilitan comprender el concepto. Desde luego, la gente no quiere concebir estas cosas. Quiere proteger los misterios hasta de un mero intento de explicación, por miedo de que una explicación pueda hacer que desaparezcan los tesoros.

Cuando a otras personas les entra la curiosidad, encuentran que «misteriosos son los caminos del Señor» es una conveniente herramienta de antipensamiento. Dando a entender que la persona que pregunta es arrogante y ambiciosas, pueden extinguir la curiosidad en un instante. Solía funcionar, y todavía funciona en comunidades donde la ignorancia de la ciencia se considera un defecto insignificante si no es que, de hecho, una virtud. Creo que deberíamos dejar de tratar esta observación «piadosa» como cualquier clase de sabiduría y reconocerla como la propaganda a la defensiva que a todas luces es. Una respuesta constructiva podría ser: «¡Qué bien! Me encantan los misterios. Veamos si también éste podemos resolverlo. ¿Se te ocurre alguna idea?»

Concebir algo nuevo es un trabajo duro, no sólo cuestión de formular alguna idea en tu mente, echarle un vistazo rapidísimo y luego refrendarla. Lo que ahora es inconcebible para nosotros puede resultar ser evidentemente concebible cuando hayamos puesto un poco más de empeño. Y cuando declaramos con seguridad que algunas cosas son verdaderamente imposibles (el número primo más grande, o un triángulo plano con ángulos interiores que sumen más que dos ángulos rectos, o un soltero casado) no es tanto porque nos parezca inconcebible como porque hemos concebido sus componentes tan bien, y tan exhaustivamente, que la imposibilidad de su conjunción es de suyo claramente concebible.

Todavía no hemos logrado concebir completamente cómo podría existir el significado en el mundo material, o cómo surgió y evolucionó la vida, o cómo trabaja la conciencia,

o si el libre albedrío puede ser uno de nuestros atributos, pero hemos hecho algún progreso: las preguntas que estamos planteando y abordando ahora son mejores que las preguntas de antaño. Estamos cerca de encontrar las respuestas.

XI. Lo que quedó fuera

Algunos lectores del borrador de este libro expresaron sorpresa y decepción porque no incluía algunas de mis mejores bombas de intuición. De hecho, varias decenas no aparecen aquí, entre ellas algunas de mis favoritas. Para algunas pienso que es necesaria una explicación.

«¿Dónde estoy?» *[Where am I?]* es probablemente mi bomba de intuición más conocida, pero por esa misma razón pudo quedar fuera. Se publicó originalmente en *Brainstorms* en 1978, y después se incluyó en 1981 en *The Mind's I.* Se han publicado traducciones en cerca de doce idiomas, y figura en varias antologías. La película *Victim of the Brain* (1984) tiene una dramatización de media hora inspirada en ella (conmigo haciendo del cuerpo del posterior Dennett), y también aparece una escena en un documental de 1981 de la BBC sobre la conciencia y el cerebro. Y luego está la dramatización para sombras chinescas javanesas que produjo la conocida marionetista Lynn Jeffries en el teatro Loeb de Harvard en 1984. Google no deja duda de que se consigue fácilmente, junto con un torrente de comentarios.

«La balada de la pizzería Shakey's» *[The Ballad of Shakey's Pizza Parlor]* hace un serio intento de desmantelar una serie de presuposiciones acerca de «la creencia *de re* y la creencia *de dicto*», que alguna vez dominó el pensamiento de algunos filósofos que estudiaban la intencionalidad, pero que al resto del mundo le resultaba poco familiar.[1] Si la hubiera incluido aquí, primero habría tenido que infectar a mis lectores con algunas intuiciones seductoras pero equivocadas, para luego poder curarlos con mi bomba de intuición. Algunos filósofos tienen que conocer bien el tema, pero otros pueden seguir felizmente ignorantes de él sin perderse de nada.

En la parte sobre evolución tuve que dejar fuera, a regañadientes, mi nueva herramienta de pensamiento favorita, los espacios darwinianos del filósofo Peter GodfreySmith, el mejor empleo de un espacio multidimensional como herramienta de pensamiento en la filosofía que yo conozca, porque para ser efectiva habría requerido un resumen demasiado largo de la teoría evolutiva y los fenómenos biológicos. Además, la explico a otros filósofos con términos un poco técnicos, y presento algunos de sus magníficos usos, en mi reseña ensayo «Homunculi Rule»,[2] en el libro de GodfreySmith *Population Thinking and Natural Selection.*[3]

Mi ensayo «Quining Qualia»[4] consta de no menos de 14 bombas de intuición

concebidas para aclarar, y luego desterrar por desesperadamente confuso, el concepto filosófico de los qualia. Sólo uno de ellos se incluye aquí, La maldición de la coliflor, para ayudar a introducir el concepto y un problema serio con los qualia. «Quining Qualia» puede considerarse una lectura complementaria para cualquiera que siga pensando que el concepto de qualia (como les gusta definirlo a los filósofos) es una buena idea. El ensayo aparece en varias antologías y se consigue fácilmente en internet, en varios idiomas. Mi libro *Dulces sueños: obstáculos filosóficos para una ciencia de la conciencia*[5] contiene todavía más argumentos y bombas de intuición sobre el tema. Otros sobre la conciencia son «curare-cum-amnestic», en «Why You Can't Make a Computer That Feels Pain»,[6] «Swamp Mary and RoboMary», en «What RoboMary Knows»,[7] y modelos de conciencia orwellianos y estalinistas en «Orwellian and Stalinesque», en *Consciousness Explained.*[8] Para todos éstos se requiere una escenografía más grande que la que podía montar cómodamente si quería que este libro fuera relativamente corto.

También he dejado fuera las diversas bombas de intuición para pensar sobre la religión que presenté en *Romper el hechizo: la religión como fenómeno natural,*[9] y mi ejemplo de Supermán alterando el Esquisto de Burgess en *Science and Religion: Are They Compatible?*[10]

[1] D. C. Dennett, «Beyond Belief», en A. Woodfield (comp.), *Thought and Object: Essays on Intentionality,* Oxford University Press, Oxford, 1982.

[2] D. C. Dennett, «*Homunculi* Rule…», *op. cit.*

[3] P. Godfrey-Smith, *Darwinian Populations and Natural Selection,* Oxford University Press, Oxford, 2009. Véase también la respuesta de Godfrey-Smith en «Agents and Acacias: Replies to Dennett, Sterelny, and Queller», *Biology and Philosophy,* 26 (4): 501-515, 2011.

[4] D. C. Dennett, «Quining Qualia», en A. J. Marcel y E. Bisiach (comps.), *Consciousness in Modern Science,* Oxford University Press, Oxford, 1988, pp. 42-77.

[5] D. C. Dennett, *Sweet Dreams, op. cit.*

[6] D. C. Dennett, «Why You can't Make a Computer that Feels Pain», *Synthese,* 38 (3): 415-456, 1978.

[7] D. C. Dennett, «What RoboMary Knows», en T. Alter y S. Walter (comps.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism,* Oxford University Press, Oxford, 2007, pp. 15-31.

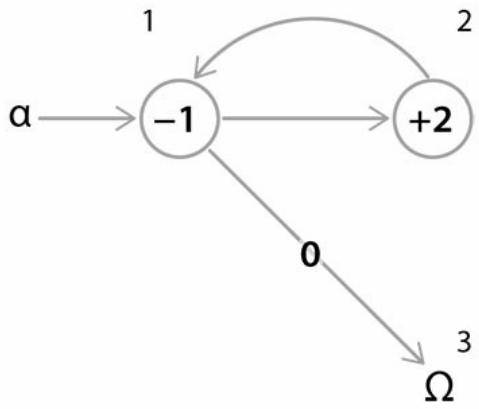
[8] D. C. Dennett, *Consciousness Explained, op. cit.*

[9] D. C. Dennett, *Breaking the Spell, op. cit.*

[10] D. C. Dennett y A. Plantinga, *Science and Religion: Are They Compatible?,* Oxford University Press, Oxford, 2011.

Apéndice
Soluciones a los problemas de la máquina de registro

EJERCICIO 1
Programa 1



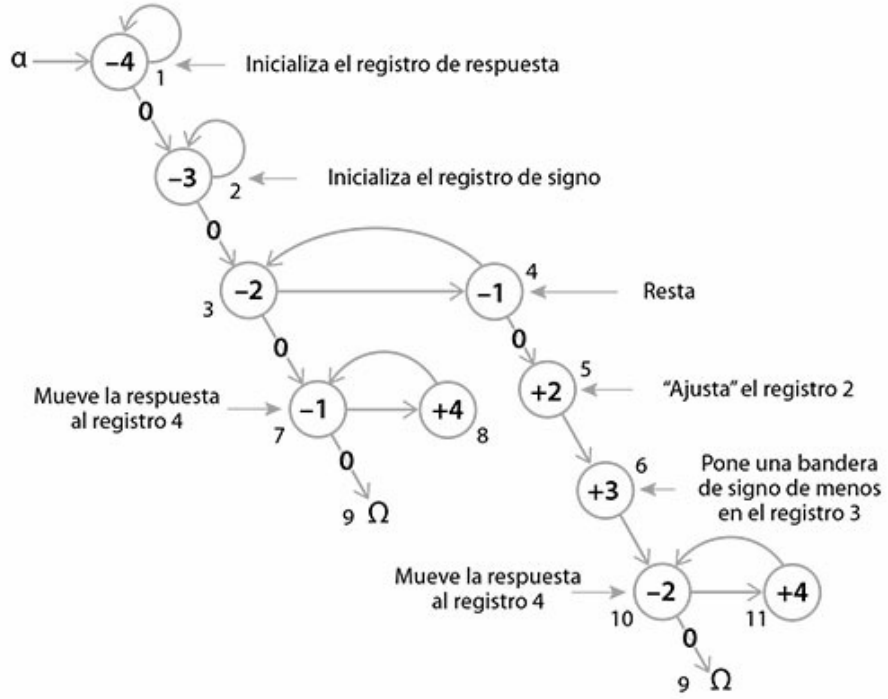
a) ¿Cuántos pasos le tomará a la máquina de registro sumar 2 + 5 y obtener 7 ejecutando el programa 1 (contando *fin* como paso)?
Respuesta: seis pasos: tres decrementos, dos incrementos, un Fin *(el último decremento es una ramificación en cero).*

b) ¿Cuántos pasos le tomará sumar 5 + 2?
Respuesta: doce pasos: seis decrementos, cinco incrementos y un Fin.

(¿Qué conclusión sacas de esto?)
Respuesta: el orden de los contenidos puede significar una gran diferencia, así que podrías pensar que sería bueno tener la regla de siempre poner el número más chico en el registro 1 —pero si antes tuvieras que probar los dos números para ver cuál es más chico, usarías más pasos que los que te tomaría hacer la suma—.

EJERCICIO 2

a) Escribe el programa PER para este diagrama de flujo. (Observa que, como el programa se ramifica, puedes numerar los pasos de varias maneras diferentes. No importa cuál escojas, siempre y cuando los comandos «ve a» apunten a los pasos correctos.)



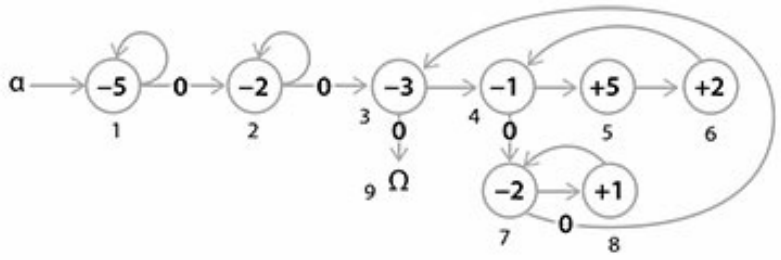
Paso	*Instrucción*	*Registro*	*Ve al paso*	*[Ramifica al paso]*
1	*Dir*	4	1	2
2	*Dir*	3	2	3
3	*Dir*	2	4	7
4	*Dir*	1	3	5
5	*Inc*	2	6	
6	*Inc*	3	10	
7	*Dir*	1	8	9
8	*Inc*	4	7	
9	*Fin*			
10	*Dir*	2	11	9
11	*Inc*	4	10	

b) ¿Qué pasa cuando el programa trata de restar 3 de 3 o 4 de 4?
El programa se detiene con 0 en el registro 4.

c) ¿Qué error posible se evita reduciendo a cero el registro 3 antes de tratar de restar en el paso 3 y no después del paso 4?
Si hubiera un cero al principio tanto en el registro 1 como en el 2, la respuesta cuando finalizara el programa podría ser absurda (ya fuera –0, o algún número distinto de 0 o 1 en el registro de signo).

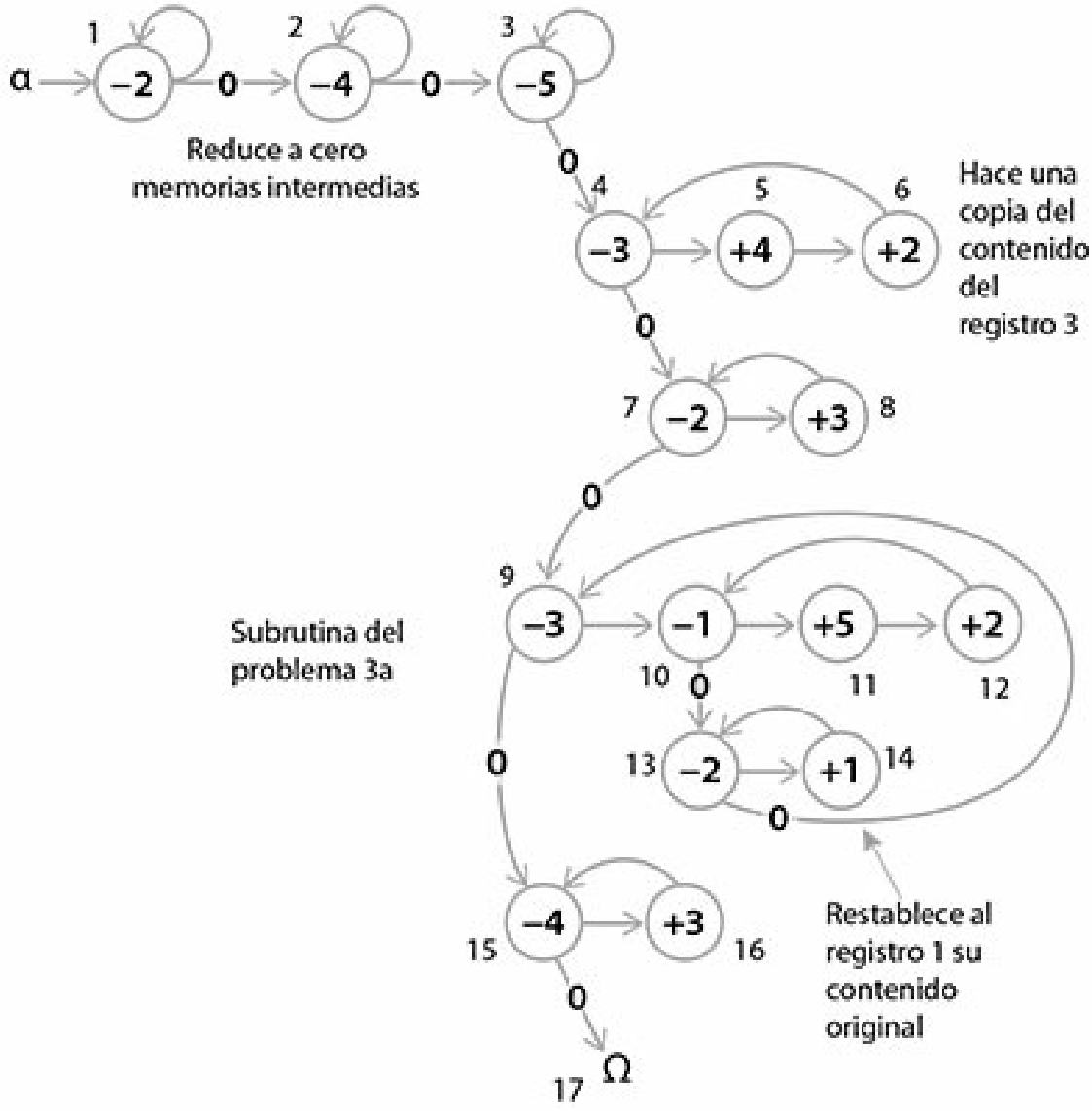
EJERCICIO 3

a) Dibuja un diagrama de flujo (y escribe el programa PAR) para multiplicar el contenido del registro 1 por el contenido del registro 3, y pon la respuesta en el registro 5.



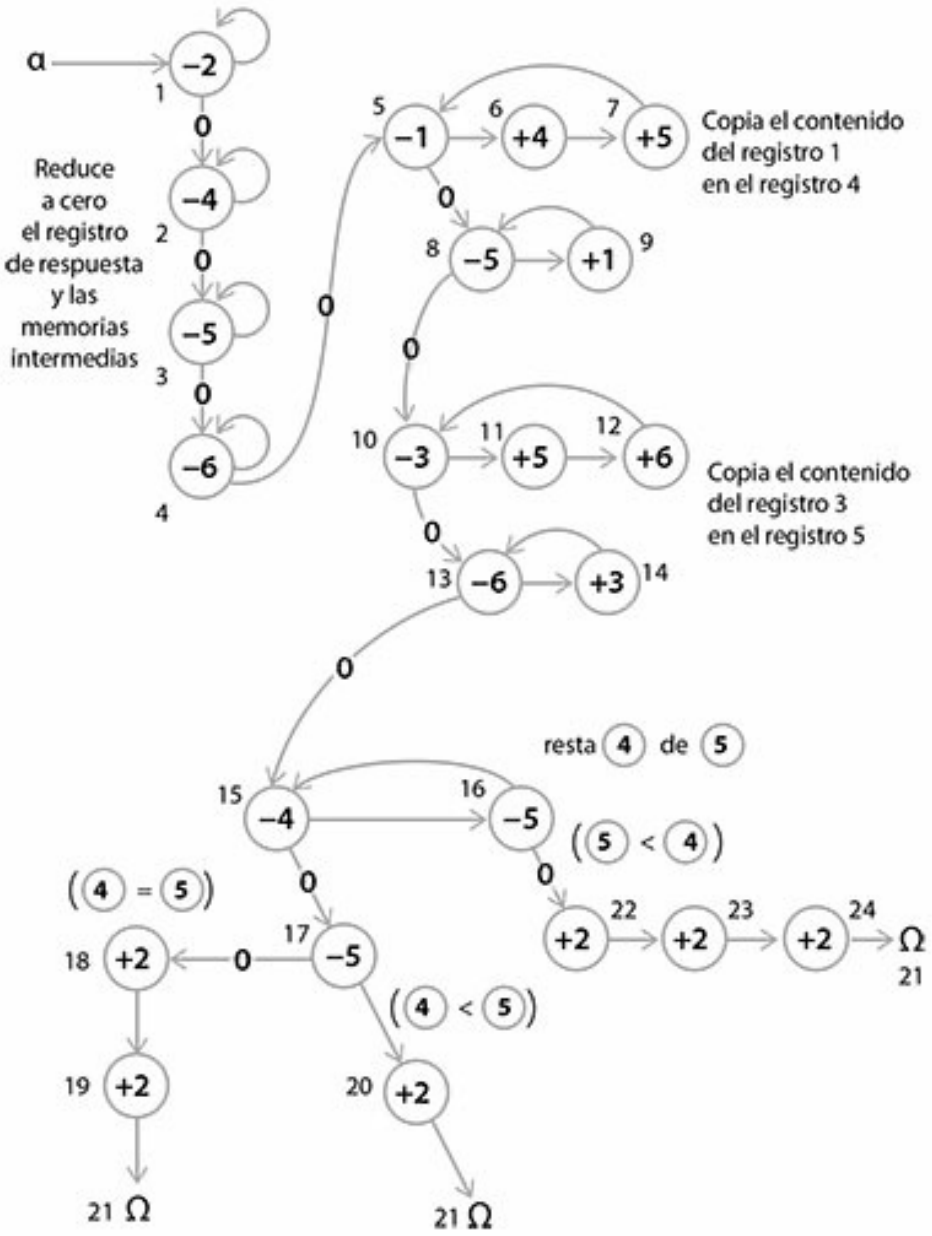
Paso	*Instrucción*	*Registro*	*Ve al paso*	*[Ramifica al paso]*
1	*Dir*	5	1	2 [reduce a cero memorias intermedias]
2	*Dir*	2	2	3
3	*Dir*	3	4	9 [comienza el conteo regresivo]
4	*Dir*	1	5	7 [suma a sí mismo el contenido del registro 1]
5	*Inc*	5	6	
6	*Inc*	2	4	
7	*Dir*	2	8	3
8	*Inc*	1	7	
9	*Fin*			

b) (Opcional.) Copiando y moviendo, mejora el multiplicador que creaste en el problema *a:* cuando se detiene, los contenidos originales de los registros 1 y 3 se restablecen, así que fácilmente puedes revisar que las entradas y salidas sean las correctas después de una ejecución.


α
1
−2
0
2
−4
0
3
−5
Reduce a cero
memorias intermedias
0
4
−3
5
+4
6
+2
Hace una
copia del
contenido
del
registro 3
0
7
−2
+3
8
0
9
−3
−1
+5
+2
10
0
11
12
Subrutina del
problema 3a
0
13
−2
+1
14
0
Restablece al
registro 1 su
contenido
original
15
−4
+3
16
0
17
Ω

Paso	*Instrucción*	*Registro*	*Ve al paso*	*[Ramifica al paso]*
1	*Dir*	2	1	2 [reduce a cero 3 memorias intermedias]
2	*Dir*	4	2	3
3	*Dir*	5	3	4
4	*Dir*	3	5	7 [hace una copia del contenido de 3]
5	*Inc*	4	6	
6	*Inc*	2	4	
7	*Dir*	2	8	9
8	*Inc*	3	7	
9	*Dir*	3	10	15 [esto sólo es el código del ejercicio 3a con diferentes números de fila]
10	*Dir*	1	11	13
11	*Inc*	5	12	
12	*Inc*	2	10	
13	*Dir*	2	14	9
14	*Inc*	1	13	
15	*Dir*	4	16	17 [restablece al registro 3 su valor original (el registro 1 se restablece en los pasos 13 y 14)]
16	*Inc*	3	15	
17	*Fin*			

c) (Opcional.) Dibuja un diagrama de flujo y escribe un programa PER que examine los contenidos de los registros 1 y 3 (¡sin destruirlos!) y escriba la dirección (1 o 3) del contenido del registro 2, que es menor, y ponga 2 en el registro 2 si los contenidos de los registros 1 y 3 son iguales (después de que este programa se haya ejecutado, los contenidos de los registros 1 y 3 no deberían haber cambiado, y el registro 2 debería decir si sus contenidos son iguales y, si no, cuál de ellos tiene el contenido que es menor).


α
−2
1
0
Reduce
a cero
el registro
de respuesta
y las
memorias
intermedias
−4
2
0
−5
3
0
−6
4
0
5
−1
6
+4
7
+5
Copia el contenido
del registro 1
en el registro 4
0
8
−5
+1
9
0
10
−3
11
+5
12
+6
Copia el contenido
del registro 3
en el registro 5
0
13
−6
+3
14
0
resta 4 de 5
15
−4
16
−5
(5 < 4)
0
0
(4 = 5)
17
−5
18
+2
0
(4 < 5)
19
+2
20
+2
21 Ω
21 Ω
22
+2
23
+2
24
+2
Ω
21

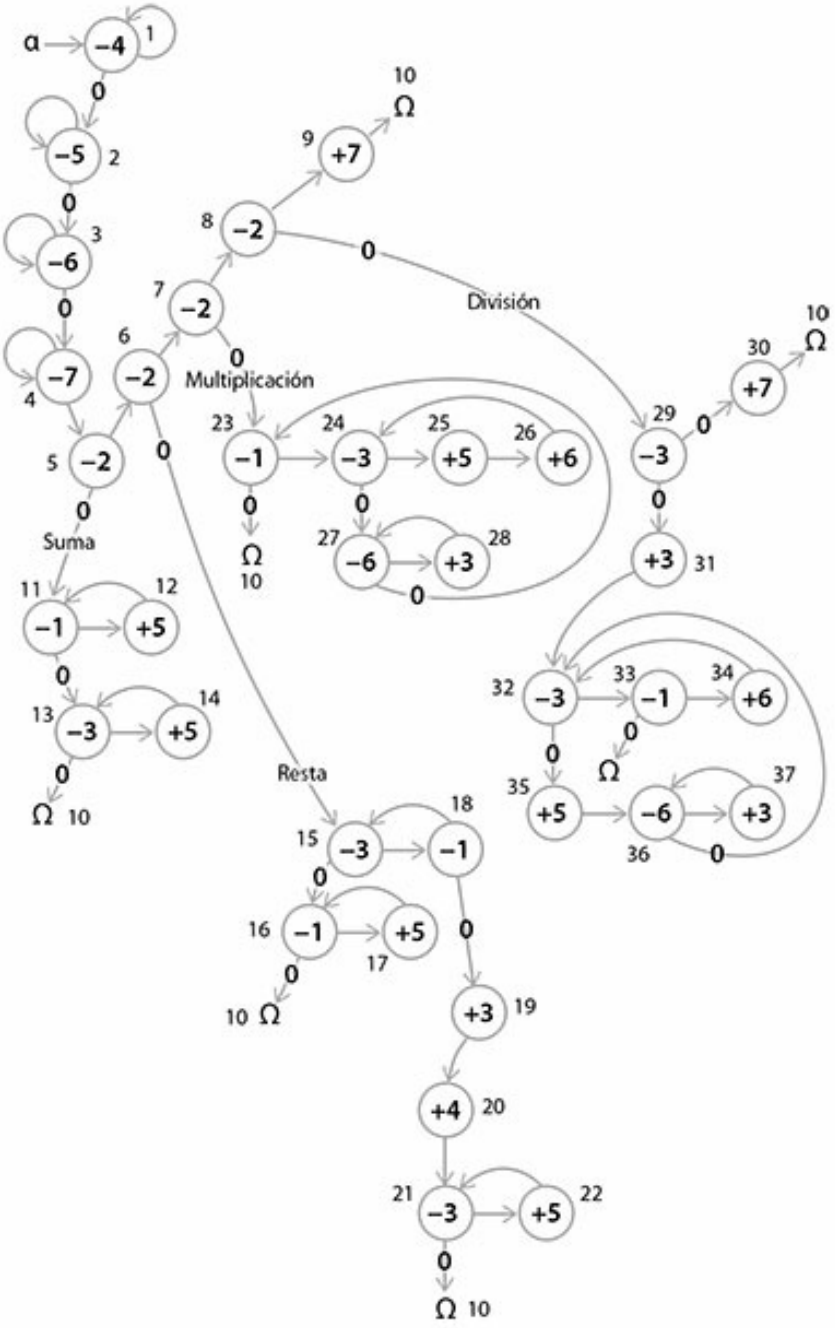
Paso	*Instrucción*	*Registro*	*Ve al paso*	*[Ramifica al paso]*	
1	*Dir*	2	1	2	[reduce a cero el registro de respuestas]
2	*Dir*	4	2	3	[reduce a cero las memorias intermedias]
3	*Dir*	5	3	4	
4	*Dir*	6	4	5	
5	*Dir*	1	6	8	[copia el contenido del registro 1 en el registro 4]
6	*Inc*	4	7		
7	*Inc*	5	5		
8	*Dir*	5	9	10	
9	*Inc*	1	8		
10	*Dir*	3	11	13	[copia el contenido del registro 3 en el registro 5]
11	*Inc*	5	12		
12	*Inc*	6	10		
13	*Dir*	6	14	15	
14	*Inc*	3	13		
15	*Dir*	4	16	17	[resta el registro 4 del registro 5]
16	*Dir*	5	15	22	
17	*Dir*	5	20	18	[revisa para verificar si el contenido de 4 = el contenido de 5]
18	*Inc*	2	19		[los contenidos son iguales; pone 2 en el registro 2]
19	*Inc*	2	21		
20	*Inc*	2	21		[el contenido de 4 es menor que el de 5; pone 1 en el registro 2]
21	*Fin*				
22	*Inc*	2	23		[el contenido de 5 es menor que el de 4; pone 3 en el registro 2]
23	*Inc*	2	24		
24	*Inc*	2	21		

EJERCICIO 4
(Opcional.)

Dibuja un diagrama de flujo y escribe un programa PER que convierta una máquina de registro en una simple calculadora de bolsillo, así:

a) Usa el registro 2 para la operación:
0 = suma
1 = resta
2 = multiplica
3 = divide

b) Pon los valores que deben seguirse en los registros 1 y 3.
(Por lo tanto, 3 0 6 significaría 3 + 6, y 5 1 3 significaría 5 − 3, y 4 2 5 significaría 4 × 5, y 9 3 3 significaría 9 ÷ 3.) Luego pon los resultados de la operación en los registros 4 a 7 y usa el registro 4 para el signo (con 0 para + y 1 para −) y el registro 5 para la respuesta numérica, el registro 6 para cualquier residuo en el caso de una división, y el registro 7 como alarma que señale un error de dividir entre cero.


α
−4
1
0
−5
2
0
−6
3
0
−7
4
−2
5
0
Suma
−2
6
−2
7
0
Multiplicación
−2
8
+7
9
10
Ω
0
División
11
−1
12
+5
0
13
−3
14
+5
0
Ω
10
0
Resta
15
−3
18
−1
0
16
−1
+5
17
0
10
Ω
0
+3
19
+4
20
21
−3
+5
22
0
Ω
10
23
−1
24
−3
25
+5
26
+6
0
Ω
10
0
27
−6
+3
28
0
29
−3
0
+7
30
10
Ω
0
+3
31
32
−3
33
−1
34
+6
0
Ω
0
35
+5
−6
+3
37
36
0

Paso	*Instrucción*	*Registro*	*Ve al paso*	*Ramifica al paso*
1	*Dir*	4	1	2 [reduce a cero registros de respuesta]
2	*Dir*	5	2	3
3	*Dir*	6	3	4
4	*Dir*	7	4	5
5	*Dir*	2	6	11 [ramifica a suma]
6	*Dir*	2	7	15 [ramifica a resta]
7	*Dir*	2	8	23 [ramifica a multiplicación]
8	*Dir*	2	9	28 [ramifica a división]
9	*Inc*	7	10	[pone bandera de código de operación inapropiada ingresada en el registro 2 y se detiene]
10	*Fin*			
11	*Dir*	1	12	13 [comienza suma]
12	*Inc*	5	11	
13	*Dir*	3	14	10
14	*Inc*	5	13	
15	*Dir*	3	18	16
16	*Dir*	1	17	10 [empieza y termina resta]
17	*Inc*	5	16	
18	*Dir*	1	15	19
19	*Inc*	3	20	[restablece registro 3]
20	*Inc*	4	21	[pone signo de menos en registro 4]
21	*Dir*	3	22	10
22	*Inc*	5	21	
23	*Dir*	1	24	10 [empieza y termina multiplicación]
24	*Dir*	3	24	27
25	*Inc*	5	26	
26	*Inc*	6	24	
27	*Dir*	6	28	23
28	*Inc*	3	27	
29	*Dir*	3	31	30 [empieza división, revisa si no hay dividir entre cero]
30	*Inc*	7	10	[pone bandera de error de dividir entre cero y se detiene]
31	*Inc*	3	32	[restablece registro 3]
32	*Dir*	3	33	35
33	*Dir*	1	34	10 [termina división]
34	*Inc*	6	32	
35	*Inc*	5	36	
36	*Dir*	6	37	32
37	*Inc*	3	36	

Fuentes documentales

I. Introducción

«El carcelero caprichoso» y «Las joyas en el bote de basura» están sacadas, con revisiones, de *Elbow Room: The Varieties of Free Will Worth Wanting* , The MIT Press, Cambridge, 1984.

II. Doce herramientas generales del pensamiento

1. «Cometer errores» («Making Mistakes») empezó como «How to Make Mistakes», en J. Brockman y K. Matson (comps.), *How Things Are,* William Morrow, Nueva York, 1995, pp. 137-144.
2. «Por parodia de razonamiento» («By Parody of Reasoning») no ha sido publicada hasta ahora.
3. «Las reglas de Rapoport» («Rapoport's Rules»). La primera vez que presenté las reglas en una publicación fue en mis reflexiones sobre *The God Delusion* de Richard Dawkins, en *Free Inquiry,* 27 (1): 64, 2007.
4. «La ley de Sturgeon» («Sturgeon's Law»). Se discute la ley en «Holding a Mirror up to Dupré», *Philosophy and Phenomenological Research,* 69 (2): 473-483, 2004.
5. «La navaja de Ockham» («Occam's Razor») no ha sido publicada hasta ahora.
6. «La escoba de Ockham» («Occam's Broom»). En muchos sitios web se atribuye la acuñación a Sydney Brenner. Al parecer, nunca ha publicado nada sobre ella.
7. «Usar público lego como señuelo» («Using Lay Audiences as Decoys») no ha sido publicada hasta ahora.
8. «Pensar afuera de la caja» («Jootsing») incluye material tomado de «I Could not Have Done Otherwise— So What?», *Journal of Philosophy,* 81 (10): 553-565, 1984.
9. «Tres especies de *gouldeo*» («Three Species of *Goulding*») está tomado de «Confusion over Evolution: An Exchange», *The New York Review of Books,* 40 (1-2): 43-44, 1993, y de «Shall We Tango? No, but Thanks for Asking», *Journal of Consciousness Studies,* 18 (5-6): 23-34, 2011.
10. «El operador "No cabe duda"» («The "Surely" Operator») nació de observaciones a «Get Real», *Philosophical Topics,* 22: 505-556, 1994.
11. «Preguntas retóricas» («Rhetorical Questions») no ha sido publicada hasta ahora.
12. «¿Qué es una profulidad? («What Is a Deepity?») recurre a material de «With a Little Help from My Friends», en D. Ross, A. Brook y D. Thompson (comps.), *Dennett's Philosophy: A Comprehensive Assessment,* The MIT Press, Cambridge, 2000, pp. 327-388.

III. Herramientas para pensar acerca del significado o contenido

13. «Asesinato en la Plaza de Trafalgar» («Murder in Trafalgar Square») está tomado de «Three Kinds of Intentional Psychology», en Richard Healey (comp.), *Reduction, Time and Reality,* Cambridge University Press, Cambridge, 1981, pp. 37-61.
14. «Un hermano mayor que vive en Cleveland» («An Older Brother Living in Cleveland») está tomado de «Brain Writing and Mind Reading», en Keith Gunderson (comp.), *Language, Mind and Knowledge. Minnesota Studies in the Philosophy of Science,* vol. 7, University of Minnesota Press, Minneapolis, 1975, pp. 403-416.
15. «Mi papá es doctor» («"Daddy Is a Doctor"») está tomado de *Content and Consciousness,* Routledge & Kegan Paul, Londres, 1969.
16. «Imagen manifiesta e imagen científica» («Manifest Image and Scientific Image») incluye material tomado de «Expecting Ourselves to Expect», *Behavioral and Brain Sciences,* 36(3): 209-210, 2013; «Sakes and Dints», *The Times Literary Supplement,* 2: 12-14, 2012, y «Kinds of Things», en D. Ross, J. Ladyman y H. Kincaid (comps.), *Scientific Metaphysics,* Oxford University Press, Oxford, 2013, pp. 96-107. La discusión sobre los osos hormigueros y los pájaros está tomada de *Elbow Room: The Varieties of Free Will Worth Wanting ,* The MIT Press, Cambridge, 1984.
17. «Psicología vulgar» («Folk Psychology») está tomada de «Three Kinds of Intentional Psychology», en Richard Healey (comp.), *Reduction, Time and Reality,* Cambridge University Press, Cambridge, 1981, pp. 37-61.
18. «La postura intencional» («The Intentional Stance») está tomada de «Intentional Systems», *Journal of Philosophy,* 68: 87-106, 1971, y de *The Intentional Stance,* The MIT Press, Cambridge, 1987.
19. «La distinción personal/subpersonal» («The Personal/Sub-personal Distinction») está tomada de *Content and Consciousness,* Routledge & Kegan Paul, Londres, 1969.
20. «Una cascada de homúnculos» («A Cascade of Homunculi») recurre a *Brainstorms,* The MIT Press, Cambridge, 1978, y a «2008: What have You Changed Your Mind About? Why? Competition in the Brain», <Edge.org>, The World Question Center, [s. l.], 2008, que posteriormente se publicó en John Brockman (comp.), *What Have You Changed Your Mind About? Today's Leading Minds Rethink Everything ,* Harper Collins, Nueva York, 2009.
21. «El operador *maomeno*» («The *Sorta* Operator») no ha sido publicada hasta ahora, pero unos temas que ahí se tocan se amplían en «Turing's "Strange Inversion of Reasoning"», en S. B. Cooper y Jan van Leeuwen (comps.), *Alan Turing. His Work and Impact,* Elsevier, Nueva York, 2013, pp. 569-573.
22. «Tejido maravilloso» («Wonder Tissue»). Usé el término por primera vez en una publicación en «Natural Freedom», *Metaphilosophy,* 36 (4): 449-459, 2005. Agradezco

a Christof Koch por hacerme reparar en la cita de William Bateson.
23. «Atrapado en el centro de operaciones del robot» («Trapped in the Robot Control Room») está tomada de «Current Issues in the Philosophy of Mind», *American Philosophical Quarterly,* 15 (4): 249-261, 1978.

24. «Los siete secretos del poder de la computación revelados» («The Seven Secrets of Computer Power Revealed») surgió de un curso introductorio a las ciencias de la computación que George Smith, David Isles y yo impartimos en Tufts, con la extraordinaria ayuda del estudiante de licenciatura con especialización en ciencias de la computación Steve Barney. Steve se convirtió en el programador principal en el Estudio de *Software* Curricular en Tufts, proyecto financiado por la Fundación Sloan Foundation tras ver la computadora animada, aesop, que Steve creó para ese curso. RodRego se inspiró en aesop, y lo programé en Logo en 1986 (lo llamé Rego y la figura del homúnculo de cabeza de pico en el apartado 24 es una huella fósil de la «tortuga» de Logo). Luego se volvió a hacer en versiones más portátiles y resistentes; el primero en hacerlo fue Rod Da Silva —RodRego— y luego lo revisó Nikolai Shvertner en el Estudio de *Software* Curricular.

25. «Máquinas virtuales» («Virtual Machines»). He discutido estas máquinas en muchos lugares, entre ellos «Notes on Prosthetic Imagination», *The New Boston Review,* 7 (3): 3-7, 1982, el manifiesto que llevó a la creación del Estudio de *Software* Curricular; *Consciousness Explained,* Little Brown, Boston, 1991, y «The Practical Requirements for Making a Conscious Robot», *Proceedings of the Royal Society, A,* 349: 133-146, 1994.

26. «Algoritmos» («Algorithms») recurre a material de *Darwin's Dangerous Idea: Evolution and the Meanings of Life,* Simon & Schuster, Nueva York, 1995.

27. «Automatizar el elevador» («Automating the Elevator») lo presenté en «Varieties of Content», Presentation at Concepts: Content and Constitution, un simposio llevado a cabo en la Universidad de Copenhague el 12 de mayo de 2007. No ha sido publicada. El resumen incluye ideas de «Turing's "Strange Inversion of Reasoning"», en S. B. Cooper y J. van Leeuwen (comps.), *Alan Turing. His Work and Impact,* Elsevier, Nueva York, 2013, pp. 569-573.

28. «Algo con los pelirrojos» («A Thing about Redheads») está tomada de «Things about Things», en J. Branquinho (comp.), *The Foundations of Cognitive Science,* Clarendon Press, Oxford, 2001, pp. 133-149.
29. «El tragacuartos ambulante, la Tierra Gemela y el robot gigante» está tomada, con revisiones, de «Evolution, Error and Intentionality», en D. Partridge y Y. Wilks (comps.), *Sourcebook on the Foundations of Artificial Intelligence,* University of New Mexico Press, Albuquerque, 1988, pp. 190-211.
30. «Traducción radical y un crucigrama quineano» («Radical Translation and a Quinian Crossword Puzzle») se publicó por primera vez en «With a Little Help from My Friends», en D. Ross, A. Brook y D. Thompson (comps.), *Dennett's Philosophy: A Comprehensive Assessment,* The MIT Press, Cambridge, 2000, pp. 327-388.
31. «Motores semánticos y motores sintácticos» («Semantic Engines and Syntactic Engines») se toma con revisiones de «Three Kinds of Intentional Psychology», en R. Healey (comp.), *Reduction, Time and Reality,* Cambridge University Press, Cambridge, 1981, pp. 37-61.
32. «El hombre del pantano conoce a un tiburón vaca» («Swampman Meets a Cow-Shark») incluye material tomado, con revisiones, de «Features of Intentional Action», *Philosophy and Phenomenological Research,* 29: 232-244, 1968, y «Get Real», *Philosophical Topics,* 22: 505-556, 1994.
33. «Dos cajas negras» («Two Black Boxes») está tomado, con revisiones, de *Darwin's Dangerous Idea: Evolution and the Meanings of Life,* Simon & Schuster, Nueva York, 1995.

VI. Herramientas para pensar acerca de la evolución

34. «Ácido universal» («Universal Acid») está tomado, con revisiones, de *Darwin's Dangerous Idea: Evolution and the Meanings of Life,* Simon & Schuster, Nueva York, 1995.
35. «La biblioteca de Mendel: *Extremada* y *Evanescente*» («The Library of Mendel: *Vast* and *Vanishing*») está tomada, con revisiones, de *Darwin's Dangerous Idea: Evolution and the Meanings of Life,* Simon & Schuster, Nueva York, 1995.
36. «Los genes como palabras o como subrutinas» («Genes as Words or as Subroutines») es nueva, inspirada en R. Dawkins, *The Ancestor's Tale: A Pilgrimage to the Dawn of Time,* Weidenfeld & Nicolson, Londres, 2004, e incluye una larga cita de ese libro.
37. «El árbol de la vida» («The Tree of Life») es nueva; la figura es creación de Leonard Eisenberg. En su sitio web, <http://evogeneao.com/tree.html>, encontrarás lúcidas explicaciones de cómo usar el diagrama, y se pueden encargar carteles, playeras, sudaderas y otros artículos.
38. «Grúas y ganchos celestes: levantamiento en el espacio del diseño» («Cranes and Skyhooks, Lifting in Design Space») se toma, con revisiones, de *Darwin's Dangerous Idea: Evolution and the Meanings of Life,* Simon & Schuster, Nueva York, 1995.
39. «Competencia sin comprensión» («Competence without Comprehension») está tomada, con revisiones, de «Darwin's "Strange Inversion of Reasoning"», *Proceedings of the National Academy of the Sciences of the United States of America,* 106 (supl. 1): 10061-10065, 2009.
40. «Razones de libre flotación» («Free-Floating Rationales»). Las razones de libre flotación se presentaron en «Intentional Systems in Cognitive Ethology: The "Panglossian Paradigm" Defended», *Behavioral and Brain Sciences,* 6 (3): 343-390, 1983, y se han discutido en muchos otros ensayos.
41. «¿Entienden los chapulines los números primos?» («Do Locusts Understand Prime Numbers?») está tomada, con revisiones, de *Darwin's Dangerous Idea: Evolution and the Meanings of Life,* Simon & Schuster, Nueva York, 1995.
42. «Cómo explicar los saltos de las gacelas» («How to Explain Stotting») es nueva.
43. «Cuidado con el Mamífero Primario» («Beware of the Prime Mammal») está tomada, con revisiones, de *Darwin's Dangerous Idea: Evolution and the Meanings of Life,* Simon & Schuster, Nueva York, 1995.
44. «¿Cuándo ocurre la especiación?» («When Does Speciation Occur?») está tomada, con revisiones, de D. Dennett y A. Kathleen, «The Multiple Drafts Model», *Scholarpedia,* 3 (4): 4321, 2008.
45. «Hacedores de viudas, Eva Mitocondrial y coronaciones retroactivas»

(«Widowmakers, Mitochondrial Eve, and Retrospective Coronations») está tomada, con revisiones, de *Darwin's Dangerous Idea: Evolution and the Meanings of Life,* Simon & Schuster, Nueva York, 1995.
46. «Ciclos» («Cycles») está tomada, con revisiones, de «Cycles», <Edge.org>, The World Question Center, [s. l.], 2011, mi respuesta a la Edge Question 2011 «What Scientific Concept Would Improve Everybody's Cognitive Toolkit?», que más adelante se publicó en John Brockman (comp.), *This Will Make You Smarter. New Scientific Concepts to Improve Your Thinking ,* Harper Collins, Nueva York, 2012.
47. «¿Qué *le dice* el ojo de la rana al cerebro de la rana? («What *Does* the Frog's Eye Tell the Frog's Brain?») está tomada, con revisiones, de *Darwin's Dangerous Idea: Evolution and the Meanings of Life,* Simon & Schuster, Nueva York, 1995.
48. «Saltar por el espacio en la biblioteca de Babel» («Leaping through Space in the Library of Babel») está tomada, con revisiones, de *Darwin's Dangerous Idea: Evolution and the Meanings of Life,* Simon & Schuster, Nueva York, 1995.
49. «¿Quién es el autor de *Spamlet*?» («Who Is the Author of *Spamlet*?) está tomada, con revisiones, de mi alocución presidencial a la división este de la American Philosophical Association, titulada «In Darwin's Wake, Where Am I?», *Proceedings and Addresses of the American Philosophical Association,* 75 (2): 13-30, 2001.
50. «Ruido en el hotel virtual» («Noise in the Virtual Hotel») está tomada, con revisiones, de «Collision-Detection, Muselot, and Scribble: Some Reflections on Creativity», en D. Cope, *Virtual Music, Computer Synthesis of Musical Style,* The MIT Press, Cambridge, 2001, pp. 283-291.
51. «Herb, Alice y Hal, el bebé» («Herb, Alice, and Hal, the Baby») está tomada, con revisiones, de «Homunculi Rule: *Reflections on Darwinian Populations and Natural Selection* by Peter Godfrey-Smith, Oxford University Press, 2009», *Biology and Philosophy,* 26 (4): 475-488, 2010.
52. «Memes» («Memes») recurre a diversas publicaciones sobre el tema mencionadas en el ensayo.

53. «Dos contraimágenes» («Two Counter-images») incluye material tomado de *Consciousness Explained,* Little Brown, Boston, 1991, y *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness,* The MIT Press, Cambridge, 2005.
54. «La corazonada zombi» («The Zombie Hunch»). La corazonada zombi se definió en gran parte en *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness,* The MIT Press, Cambridge, 2005.
55. «Zombis y zimbos» («Zombies and Zimboes») recurre a revisiones de material publicado en *Consciousness Explained,* Little Brown, Boston, 1991.
56. «La maldición de la coliflor» («The Curse of the Cauliflower») recurre a material revisado de «Quining Qualia», en Anthony J. Marcel y E. Bisiach (comps.), *Consciousness in Modern Science,* Oxford University Press, Oxford, 1988, pp. 42-77.
57. «Vim: ¿cuánto cuesta en "dinero de verdad?"» («Vim: How Much Is That in "Real Money "»?) está tomada, con revisiones, de «Consciousness: "How Much Is That in Real Money?"», en R. L. Gregory (comp.), *The Oxford Companion to the Mind,* 2a ed., Oxford University Press, Oxford, 2001.
58. «El triste caso del señor Clapgras» («The Sad Case of Mr. Clapgras») está tomada, con revisiones, de *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness,* The MIT Press, Cambridge, 2005.
59. «La baraja afinada» («The Tuned Deck») está tomada, con revisiones, de «Explaining the "Magic" of Consciousness», en *Exploring Consciousness, Humanities, Natural Science, Religion: Proceedings of the International Symposium, November 19-20, 2001,* Fondazione Carlo Elba, Milán, 2002, pp. 47-58.
60. «La habitación china» («The Chinese Room») recurre a material revisado de D. Hofstadter y D. C. Dennett (comps.), *The Mind's I. Fantasies and Reflections on Self and Soul,* Basic Books, Nueva York, 1981, y *The Intentional Stance,* The MIT Press, Cambridge, 1987.
61. «La caída del Teleclón de Marte a la Tierra» («The Teleclone Fall from Mars to Earth») está tomada, con revisiones, de D. Hofstadter y D. C. Dennett (comps.), *The Mind's I. Fantasies and Reflections on Self and Soul,* Basic Books, Nueva York, 1981.
62. «El yo como centro de gravedad narrativa» («The Self as the Center of Narrative Gravity»). Esta herramienta se describió en «Why Everyone is a Novelist», *The Times Literary Supplement,* 4 (16-22): 459, 1988.
63. «Heterofenomenología» («Heterophenomenology»). La heterofenomenología se describió en «How to Study Consciousness Empirically: Or Nothing Comes to Mind», *Synthese,* 53 (2): 159-180, 1982. Luego se amplió en un capítulo de *Consciousness Explained,* Little Brown, Boston, 1991. Un número especial de *Phenomenology and Cognitive Science,* con Alva Noë como directora invitada, se dedicó a la

heterofenomenología, y consta de la mayoría de las reflexiones más interesantes que después ha habido sobre el tema. Incluye mi réplica a los ensayos: «Heterophenomenology Reconsidered», *Phenomenology and Cognitive Science,* 6 (1-2): 247-270, 2007.

64. «Mary, la científica del color: una muleta de razonamiento al descubierto» («Mary the Color Scientist: A Boom Crutch Unveiled») está tomada, con revisiones, de *Consciousness Explained,* Little Brown, Boston, 1991, y «What RoboMary Knows», en T. Alter y S. Walter (comps.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism,* Oxford University Press, Oxford, 2007, pp. 15-31, que apareció en la segunda de las dos antologías dedicadas a «Mariology»; la primera de ellas es P. Ludlow, Y. Nagasawa y D. Stoljar (comps.), *There's Something about Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument,* The MIT Press, Cambridge, 2004.

VIII. Herramientas para pensar acerca del libre albedrío

65. «Una neurocirujana verdaderamente nefanda» («A Truly Nefarious Neurosurgeon») está tomada de mi conferencia Erasmus, «Erasmus: Sometimes a Spin-Doctor Is Right», en 2012.
66. «Un juguete determinista: el juego de la vida de Conway» («A Deterministic Toy: Conway's Game of Life»). Martin Gardner presentó el juego de la vida (Life) a un amplio público en dos de sus columnas «Mathematical Games» en *Scientific American,* en octubre de 1970 y febrero de 1971. William Poundstone, en *The Recursive Universe: Cosmic Complexity and the Limits of Scientific Knowledge,* William Morrow, Nueva York, 1985, ofrece una magnífica exploración del juego y sus implicaciones filosóficas. También hay un artículo de Wikipedia decoroso sobre él, con muchas ligas. A menudo he usado el juego de la vida en mis escritos. Esta versión está tomada, con revisiones, de la explicación presentada en *Freedom Evolves,* Viking Penguin, Nueva York, 2003.
67. «Piedra, papel o tijeras» («Rock, Paper, and Scissors») no ha sido publicada.
68. «Dos loterías» («Two Lotteries») está tomada, con revisiones, de *Elbow Room: The Varieties of Free Will Worth Wanting ,* The MIT Press, Cambridge, 1984.
69. «Hechos históricos inertes» («Inert Historical Facts») está tomada, con revisiones, de «Get Real», *Philosophical Topics,* 22: 505-556, 1994. Tanmoy Bhattacharya me hizo reparar en las complicaciones cuánticas.
70. «Un maratón de ajedrez por computadora»(«A Computer Chess Marathon») está tomado de *Freedom Evolves,* Viking Penguin, Nueva York, 2003.
71. «Máxima responsabilidad» («Ultimate Responsibility») no ha sido publicada hasta ahora.
72. «Sphexismo» («Sphexishness») no ha sido publicada hasta ahora.
73. «Los niños del Brasil: otra muleta de razonamiento» («The Boys from Brazil: Another Boom Crutch») no ha sido publicada hasta ahora.

IX: ¿Cómo es ser un filósofo?

74. «Un trato faustiano» («A Faustian Bargain») está tomada, con revisiones, de «Review of *Other Minds: Critical Essays, 1969-1994* by Thomas Nagel, 1995», *Journal of Philosophy,* 63 (8): 425-428, 1996.
75. «La filosofía como una autoantropología naïf» («Philosophy as Naïve Auto-anthropology») está tomada, con revisiones, de «Sakes and Dints», *The Times Literary Supplement,* 2: 12-14, 2012, y «Kinds of Things», en D. Ross, J. Ladyman y H. Kincaid (comps.), *Scientific Metaphysics,* Oxford University Press, Oxford, 2013, pp. 96-107.
76. «Verdades de orden superior sobre el *ajetrez*» («Higher-Order Truths of Chmess») está tomada, con revisiones, de «The Higher-Order Truths about Chmess», *Topoi,* 25 (1-2): 39-41, 2006.
77. «El diez por ciento que está bien» («The 10 Percent That's Good») no ha sido publicada hasta ahora.

Bibliografía

Abbott, Edwin A., *Flatland: A Romance in Many Dimensions,* 5ª ed., pról. de Isaac Asimov, Harper Collins, Nueva York, 1983 [1884]. [Versión en español: *Planilandia,* Laertes, Barcelona, 1993.]

Alain, Chartier, *Propos d'un Normand 1906-1914,* Gallimard, París, 1956 [1908]. Citado en Deborah S. Rogers y Paul R. Ehrlich, «Natural Selection and Cultural Rates of Change», *Proceedings of the National Academy of Sciences,* 105 (9): 3416-3420, 2008.

Austin, John L., *Philosophical Papers,* Oxford University Press, Oxford, 1961. [Versión en español: *Ensayos filosóficos,* trad. de Alfonso García Suárez, Revista de Occidente, Madrid, 1975.]

Axelrod, Robert, *The Evolution of Cooperation,* Basic Books, Nueva York, 1984.

———, y William Hamilton, «The Evolution of Cooperation», *Science,* 211: 1390-1396, 1981.

Barrett, Justin, «Exploring the Natural Foundations of Religion», *Trends in Cognitive Science,* 4: 29-34, 2000.

Bateson, William, «Review of the *Mechanisms of Mendelian Heredity* by T. H. Morgan, A. H. Sturtevant, H. J. Muller and C. B. Bridges, New York, 1915», *Science,* 44: 536-543, 1916.

Behe, Michael J., *Darwin's Black Box: The Biochemical Challenge to Evolution,* Free Press, Nueva York, 1996.

Bennett, Max, Daniel C. Dennett, Peter M. S. Hacker y John Searle, *Neuroscience and Philosophy: Brain, Mind, and Language,* Columbia University Press, Nueva York, 2007. [Versión en español: *La naturaleza de la conciencia: cerebro, mente y lenguaje,* trad. de Roc Filella, Paidós, Barcelona y México, 2008.]

———, y Peter M. S. Hacker, *Philosophical Foundations of Neuroscience,* Wiley-Blackwell, Malden, 2003.

Block, Ned, «Troubles with Functionalism», en W. Savage (comp.), *Perception and Cognition: Issues in the Foundations of Psychology,* vol. 9, University of Minnesota Press, Minneapolis, 1978, pp. 261-326.

———, «What is Dennett's Theory a Theory Of ?», *Philosophical Topics,* 22: 23-40, 1994. (Número especial acerca de la filosofía de Daniel C. Dennett.)

Bobrow, Daniel, *Qualitative Reasoning about Physical Systems,* The MIT Press, Cambridge [Estados Unidos], 1985.

Borges, Jorge Luis, *Ficciones,* Alianza, Madrid, 1997.

Bray, Dennis, *Wetware: A Computer in Every Living Cell,* Yale University Press, New Haven, 2009.

Brockman, John (comp.), *What Have You Changed Your Mind About? Today's Leading Minds Rethink Everything ,* HarperCollins, Nueva York, 2009.

———, *This Will Make You Smarter. New Scientific Concepts to Improve Your Thinking ,* Harper Collins, Nueva York, 2012.

Brooks, Rodney A., *Planning is Just a Way of Avoiding Figuring out What to Do Next,* MIT Artificial Intelligence Laboratory, Cambridge [Estados Unidos], 1987. Disponible en <http://people. csail.mit.edu/>.

———, «Intelligence without Representation», *Artificial Intelligence,* 47: 139-159, 1991.

Cann, Rebecca L., Mark Stoneking y Allan C. Wilson, «Mitochondrial dna and Human Evolution», *Nature,* 325: 31-36, 1987.

Chalmers, David, «Facing Up to the Problem of Consciousness», *Journal of Consciousness Studies,* 2 (3): 200-219, 1995.

Clark, Andy, «Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science», *Behavioral and Brain Sciences,* 36 (3): 181-204, 2013.

Clegg, Liam, *Protean Free Will,* California Institute of Technology, Pasadena, 2012. Disponible en <http://authors.library.caltech.edu/29887/>.

Cope, David, *The Algorithmic Composer,* A-R Editions, Middleton, 2000.

———, *Virtual Music: Computer Synthesis of Musical Style,* The MIT Press, Cambridge [Estados Unidos], 2001.

Cronin, Helena, *The Ant and the Peacock,* Cambridge University Press, Cambridge [Estados Unidos], 1991. [Versión en español: *La hormiga y el pavo real. El altruismo y la selección sexual desde Darwin hasta hoy,* trad. de Eva Zimerman de Aguirre, Norma, México y Barcelona, 1995.]

Damiaan, Denys, Mariska Mantione, Martijn Figee, Pepijn van den Munckhof, Frank Koerselman, Herman Westenberg, Andries Bosch y Rick Schuurman, «Deep Brain Stimulation of the *Nucleus accumbens* for Treatment-Refractory Obsessive-Compulsive Disorder», *Archives of General Psychiatry,* 67 (10): 1061-1068, 2010.

Davidson, Donald, «Knowing One's Own Mind», *Proceedings and Addresses of the American Philosophical Association,* 60: 441-458, 1987. Reimpreso en Donald Davidson, *Subjective, Intersubjective, Objective,* Oxford University Press, Nueva York, 2001, pp. 15-38. [Versión en español: «Conocer nuestra propia

mente», en *Subjetivo, intersubjetivo, objetivo,* trad. de Olga Fernández Prat, Cátedra, Madrid, 2003.]

Dawkins, Richard, *The Selfish Gene,* Oxford University Press, Oxford, 1976 (ed. correg. y aum., 1989). [Versión en español: *El gen egoísta: las bases biológicas de nuestra conducta,* trad. de Juana Robles Suárez y José Tola Alonso, Salvat, Barcelona, 1990.]

———, *The Extended Phenotype: The Gene as the Unit of Selection,* Oxford University Press, Oxford, 1982.

Dawkins, Richard, *The Blind Watchmaker,* Longmans, Londres, 1986. [Versión en español: *El relojero ciego,* trad. de Manuel Arroyo Fernández, RBA Editores, Barcelona, 1993.]

———, *Climbing Mount Improbable,* Viking Penguin, Londres, 1996. [Versión en español: *Escalando el Monte Improbable,* 2a ed., Tusquets, Barcelona, 2008.]

———, *The Ancestor's Tale: A Pilgrimage to the Dawn of Time,* Weidenfeld & Nicolson, Londres, 2004. [Versión en español: *El cuento del antepasado. Un viaje a los albores de la evolución,* trad. de Víctor V. Úbeda, Antoni Bosch Editor, Barcelona, 2010.]

De Vries, Peter, *The Value of Laughter,* Little Brown, Boston, 1953.

Dehaene, Stanislas, y J. Federico Marques, «Cognitive Euroscience: Scalar Variability in Price Estimation and the Cognitive Consequences of Switching to the Euro», *Quarterly Journal of Experimental Psychology,* 55: 705-731, 2002.

Dennett, Daniel C., «Features of Intentional Action», *Philosophy and Phenomenological Research,* 29: 232-244, 1968.

———, *Content and Consciousness,* Routledge & Kegan Paul, Londres, 1969. [Versión en español: *Contenido y conciencia,* trad. de José María Lebrón, Gedisa, Barcelona, 1996.]

———, «Intentional Systems», *Journal of Philosophy,* 68: 87-106, 1971.

———, «Brain Writing and Mind Reading», en Keith Gunderson (comp.), *Language, Mind and Knowledge. Minnesota Studies in the Philosophy of Science,* vol. 7, University of Minnesota Press, Minneapolis, 1975, pp. 403-416. Reimpreso en Daniel C. Dennett, *Brainstorms,* The MIT Press, Cambridge [Estados Unidos], 1978.

———, *Brainstorms,* The MIT Press, Cambridge [Estados Unidos], 1978.

———, «Current Issues in the Philosophy of Mind», *American Philosophical Quarterly,* 15 (4): 249-261, 1978.

———, «Why You Can't Make a Computer that Feels Pain», *Synthese,* 38 (3): 415-456, 1978.

———, «The Milk of Human Intentionality», *Behavioral and Brain Sciences,* 3

(3): 428-430, 1980. (Comentario acerca de Searle.)
———, «Three Kinds of Intentional Psychology», en Richard Healey (comp.), *Reduction, Time and Reality,* Cambridge University Press, Cambridge [Estados Unidos], 1981, pp. 37-61.
———, «Beyond Belief», en A. Woodfield (comp.), *Thought and Object: Essays on Intentionality,* Oxford University Press, Oxford, 1982. Reimpreso en Daniel C. Dennett, *The Intentional Stance,* The MIT Press, Cambridge [Estados Unidos], 1987.
———, «How to Study Consciousness Empirically: Or Nothing Comes to Mind», *Synthese,* 53 (2): 159-180, 1982.
———, «The Myth of the Computer: An Exchange», *The New York Review of Books,* 29 (11): 56-57, 1982. (Réplica a la reseña de John Searle acerca de *The Mind's I* .)
———, «Notes on Prosthetic Imagination», *The New Boston Review,* 7 (3): 3-7, 1982. Reimpreso en «30 Years of *Boston Review*», *Boston Review,* 30 (5): 40, 2005.
———, «Intentional Systems in Cognitive Ethology: The "Panglossian Paradigm" Defended», *Behavioral and Brain Sciences,* 6 (3): 343-390, 1983.
———, *Elbow Room: The Varieties of Free Will Worth Wanting* , The MIT Press, Cambridge [Estados Unidos], 1984. [Versión en español: *La libertad de acción. Un análisis de las exigencias del libre albedrío,* 2a ed., trad. de Gabriela Ventureira, Gedisa, Barcelona, 2000.]
Dennett, Daniel C., «I Could Not Have Done Otherwise— So What?», *Journal of Philosophy,* 81 (10): 553-565, 1984.
———, «The Logical Geography of Computational Approaches: A View from the East Pole», en Robert M. Harnish y M. Brand (comps.), *The Representation of Knowledge and Belief,* University of Arizona Press, Tucson, 1986, pp. 59-79.
———, *The Intentional Stance,* The MIT Press, Cambridge [Estados Unidos], 1987. [Versión en español: *La actitud intencional,* trad. de Daniel Zadunisky, Gedisa, Barcelona, 1991.]
———, «Evolution, Error and Intentionality», en Derek Partridge y Yorick Wilks (comps.), *Sourcebook on the Foundations of Artificial Intelligence,* University of New Mexico Press, Albuquerque, 1988, pp. 190-211.
———, «Quining Qualia», en Anthony J. Marcel y E. Bisiach (comps.), *Consciousness in Modern Science,* Oxford University Press, Oxford, 1988, pp. 42-77.
————, «Why Everyone is a Novelist», *The Times Literary Supplement,* 4 (16-22): 459, 1988.

———, «Memes and the Exploitation of Imagination», *Journal of Aesthetics and Art Criticism,* 48 (2): 127-135, 1990.
———, *Consciousness Explained,* Little Brown, Boston, 1991. [Versión en español: *La conciencia explicada,* trad. de Sergio Balari Ravera, Paidós, Barcelona, 1995.]
———, «Real Patterns», *Journal of Philosophy,* 88 (1): 27-51, 1991.
———, «Confusion over Evolution: An Exchange», *The New York Review of Books,* 40 (1-2): 43-44, 1993.
———, «Get Real», *Philosophical Topics,* 22: 505-556, 1994. (Respuesta a mis críticos; número especial acerca de la filosofía de Daniel C. Dennett.)
———, «The Practical Requirements for Making a Conscious Robot», *Proceedings of the Royal Society, A,* 349: 133-146, 1994.
———, *Darwin's Dangerous Idea: Evolution and the Meanings of Life,* Simon & Schuster, Nueva York, 1995. [Versión en español: *La peligrosa idea de Darwin: evolución y significados de la vida,* trad. de Cristóbal Pera Blanco-Morales, Galaxia Gutenberg, Barcelona, 1999.]
———, «How to Make Mistakes», en J. Brockman y K. Matson (comps.), *How Things Are,* William Morrow, Nueva York, 1995, pp. 137-144.
———, *Kinds of Minds: Towards an Understanding of Consciousness,* Basic Books, Nueva York, 1996. [Versión en español: *Tipos de mentes: hacia una comprensión de la conciencia,* trad. de Francisco Páez de la Cadena, Debate, Madrid, 2000.]
———, «Review of *Other Minds: Critical Essays, 1969-1994* by Thomas Nagel, 1995», *Journal of Philosophy,* 63 (8): 425-428, 1996.
———, «With a Little Help from My Friends», en Don Ross, Andrew Brook y David Thompson (comps.), *Dennett's Philosophy: A Comprehensive Assessment,* The MIT Press, Cambridge [Estados Unidos], 2000, pp. 327-388.
———, «Collision-Detection, Muselot, and Scribble: Some Reflections on Creativity», en David Cope, *Virtual Music, Computer Synthesis of Musical Style,* The MIT Press, Cambridge [Estados Unidos], 2001, pp. 283-291.
Dennett, Daniel C., «Consciousness: How Much Is That in Real Money?», en Richard L. Gregory (comp.), *The Oxford Companion to the Mind,* 2a ed., Oxford University Press, Oxford, 2001. [Versión en español: *Diccionario Oxford de la mente,* trad. de Irene Cifuentes de Castro *et al.,* Alianza, Madrid, 1995.]
———, «In Darwin's Wake, Where am I?», *Proceedings and Addresses of the American Philosophical Association,* 75 (2): 13-30, 2001 (alocución presidencial de la American Philosophical Association). Reimpreso en Jonathan

Hodge and Gregory Radick (comps.), *The Cambridge Companion to Darwin,* Cambridge University Press, Cambridge [Estados Unidos], 2003, pp. 357-376.
———, «Things about Things», en Joao Branquinho (comp.), *The Foundations of Cognitive Science,* Clarendon Press, Oxford, 2001, pp. 133-149.
———, «Explaining the "Magic" of Consciousness», en *Exploring Consciousness, Humanities, Natural Science, Religion: Proceedings of the International Symposium, November 19-20, 2001,* Fondazione Carlo Elba, Milán, 2002, pp. 47-58.
———, «The New Replicators», en Mark Pagel (comp.), *The Encyclopedia of Evolution,* vol. 1, Oxford University Press, Oxford, 2002, pp. e83-e92.
———, *Freedom Evolves,* Viking Penguin, Nueva York, 2003. [Versión en español: *La evolución de la libertad,* trad. de Ramón Vilà Vernis, Paidós, Barcelona, 2004.]
———, «Holding a Mirror up to Dupré», *Philosophy and Phenomenological Research,* 69 (2): 473-483, 2004. (Comentario acerca de John Dupré y su obra *Human Nature and the Limits of Science.*)
———, «Natural Freedom», *Metaphilosophy,* 36 (4): 449-459, 2005.
———, *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness,* The MIT Press, Cambridge [Estados Unidos], 2005. [Versión en español: *Dulces sueños: obstáculos filosóficos para una ciencia de la conciencia,* trad. de Julieta Barba y Silvia Jawerbaum, Katz, Buenos Aires, 2006.]
———, *Breaking the Spell: Religion as a Natural Phenomenon,* Viking Penguin, Nueva York, 2006. [Versión en español: *Romper el hechizo: la religión como fenómeno natural,* trad. de Felipe de Brigard, Katz, Buenos Aires, 2007.]
———, «From Typo to Thinko: When Evolution Graduated to Semantic Norms», en Stephen C. Levinson y Pierre Jaisson (comps.), *Evolution and Culture,* The MIT Press, Cambridge [Estados Unidos], 2006, pp. 133-145.
———, «The Higher-Order Truths about Chmess», *Topoi,* 25 (1-2): 39-41, 2006.
———, «*The God Delusion* by Richard Dawkins», *Free Inquiry,* 27 (1): 64, 2007.
———, «Heterophenomenology Reconsidered», *Phenomenology and Cognitive Science,* 6 (1-2): 247-270, 2007 (número especial acerca de heterofenomenología dirigido por Alva Noë).
———, «Varieties of Content», conferencia del simposio Concepts: Content and Constitution, Universidad de Copenhague, Amager, 12 de mayo de 2007.
Dennett, Daniel C., «What RoboMary Knows», en Torin Alter y Sven Walter (comps.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism,* Oxford University Press, Oxford, 2007, pp. 15-31.

————, «What Have You Changed Your Mind About? Why? Competition in the Brain», <Edge.org>, The World Question Center, [s. l.], 2008. Posteriormente publicado en John Brockman (comp.), *What Have You Changed Your Mind About? Today's Leading Minds Rethink Everything ,* Harper Collins, Nueva York, 2009.

———, «Darwin's "Strange Inversion of Reasoning"», *Proceedings of the National Academy of the Sciences of the United States of America,* 106 (supl. 1): 10061-10065, 2009.

———, «Heterophenomenology», en Tim Bayne, Alex Cleeremans y Patrick Wilken (comps.), *The Oxford Companion to Consciousness,* Oxford University Press, Oxford, 2009, pp. 345-346.

———, «Intentional Systems Theory», en Brian McLaughlin, Ansgar Beckermann y Sven Walter (comps.), *The Oxford Handbook of Philosophy of Mind,* Oxford University Press, Oxford, 2009, pp. 339-350.

———, «Cycles», <Edge.org>, The World Question Center, [s. l.], 2011 (como respuesta a la pregunta Edge 2011: «What Scientific Concept Would Improve Everybody's Cognitive Toolkit?»). Posteriormente publicado en John Brockman (comp.), *This Will Make You Smarter. New Scientific Concepts to Improve Your Thinking ,* HarperCollins, Nueva York, 2012, pp. 81-88. [Versión en español: "Ciclos", en *Este libro te hará más inteligente,* trad. de Tomás Fernández Aúz y Beatriz Eguibar, Paidós, Barcelona, 2012.]

———, «Homunculi Rule: *Reflections on Darwinian Populations and Natural Selection* by Peter Godfrey-Smith, Oxford University Press, 2009», *Biology and Philosophy,* 26 (4): 475-488, 2011. Disponible en <http://ase.tufts.edu>.

———, «Shall We Tango? No, but Thanks for Asking», *Journal of Consciousness Studies,* 18 (5-6): 23-34, 2011. (Comentarios acerca del libro de Evan Thompson *Mind in Life,* con réplicas. Los números 5 y 6 de la revista son una publicación especial acerca de Evan Thompson.)

———, «Sakes and Dints», *The Times Literary Supplement,* 2: 12-14, 2012.

———, «The Evolution of Reasons», en Bana Bashour y Hans D. Muller (comps.), *Contemporary Philosophical Naturalism and Its Implications,* Routledge, Nueva York, 2014, pp. 47-62.

———, «Expecting Ourselves to Expect», *Behavioral and Brain Sciences* 36(3): 209-210, 2013. (Comentario acerca de Clark.)

———, «Kinds of Things», en Don Ross, James Ladyman y Harold Kincaid (comps.), *Scientific Metaphysics,* Oxford University Press, Oxford, 2013, pp. 96-107.

———, «Turing's "Strange Inversion of Reasoning"», en S. Barry Cooper y Jan

van Leeuwen (comps.), *Alan Turing—His Work and Impact,* Elsevier, Nueva York, 2013, pp. 569-573.

———, y Kathleen Akins, «The Multiple Drafts Model», *Scholarpedia,* 3 (4): 4321, 2008.

———, y Alvin Plantinga, *Science and Religion: Are They Compatible?,* Oxford University Press, Oxford, 2011.

Dennett, Daniel C., y Chris F. Westbury, «Mining the Past to Construct the Future: Memory and Belief as Forms of Knowledge», en Daniel Schacter y Elaine Scarry (comps.), *Memory, Brain, and Belief,* Harvard University Press, Cambridge [Estados Unidos], 2000, pp. 11-32.

Denton, Michael, *Evolution: A Theory in Crisis,* Burnett Books, Londres, 1985.

Dewdney, Alexander K., *The Planiverse: Computer Contact with a Two-Dimensional World,* Poseidon Press, Nueva York, 1984.

Eigen, Manfred, *Steps Towards Life,* Oxford University Press, Oxford, 1992.

Ellis, Haydn, y Andrew Young, «Accounting for Delusional Misidentifications», *British Journal of Psychiatry,* 157 (2): 239-248, 1990.

Feynman, Richard, «*Surely You're Joking , Mr. Feynman!*»: *Adventures of a Curious Character,* W. W. Norton & Company, Nueva York, 1985. [Versión en español: *¿Está usted de broma, señor Feynman? Aventuras de un curioso personaje tal como fueron referidas a Ralph Leighton,* trad. de Luis Bou, Alianza, Madrid, 1987.]

Fitch, Tecumseh, «Nano-Intentionality: A Defense of Intrinsic Intentionality», *Biology and Philosophy,* 23 (2): 157-177, 2008.

FitzGibbon, Clare D., y John H. Fanshawe, «Stotting in Thomson's Gazelles: An Honest Signal of Condition», *Behavioral Ecology and Sociobiology,* 23 (2): 69-74, 1988.

Fodor, Jerry, *The Language of Thought,* Harvester Press, Hassocks, 1975. [Versión en español: *El lenguaje del pensamiento,* Alianza, Madrid, 1984.]

———, «Why Would Mother Nature Bother?», *The London Review of Books,* 25 (5): 17-18, 2003. (Reseña de *Freedom Evolves.*)

———, *LOT 2: The Language of Thought Revisited,* Oxford University Press, Oxford, 2008.

———, y Ernest Lepore, *Holism: A Shopper's Guide,* Blackwell, Oxford, 1992.

———, y Massimo Piatelli-Palmarini, *What Darwin Got Wrong ,* Farrar, Straus & Giroux, Nueva York, 2010.

Frankfurt, Harry, «Alternate Possibilities and Moral Responsibility», *Journal of Philosophy,* 65: 829-833, 1969.

Gell-Man, Murray, *The Quark and the Jaguar: Adventures in the Simple and the*

Complex, St. Martin's, Nueva York, 1995. [Versión en español: *El quark y el jaguar. Aventuras en lo simple y lo complejo,* 6a ed., trad. de Ambrosio García y Romualdo Pastor, Tusquets, Barcelona, 2007.]

Gibson, James J., *The Ecological Approach to Visual Perception,* Houghton Mifflin, Boston, 1979.

Godfrey-Smith, Peter, *Darwinian Populations and Natural Selection,* Oxford University Press, Oxford, 2009.

———, «Agents and Acacias: Replies to Dennett, Sterelny, and Queller», *Biology and Philosophy,* 26 (4): 501-515, 2011. (Réplicas acerca de las reseñas de *Darwinian Populations and Natural Selection.*)

Goffman, Erving, *The Presentation of Self in Everyday Life,* University of Edinburgh-Social Sciences Research Centre, Edimburgo, 1956. [Versión en español: *La presentación de la persona en la vida cotidiana,* trad. de Hildegarde B. Torres Perrén y Flora Setaro, Amorrortu, Buenos Aires, 1993.]

Gould, Stephen J., *Ever since Darwin,* W. W. Norton & Company, Nueva York, 1977. [Versión en español: Desde Darwin: reflexiones sobre historia natural, trad. de Antonio Resines, Crítica, Barcelona, 2010.]

Gould, Stephen J., «Tires to Sandals», *Natural History,* 98 (4): 8-15, 1989.

———, *Wonderful Life: The Burgess Shale and the Nature of History,* W. W. Norton & Company, Nueva York, 1989. [Versión en español: *La vida maravillosa,* Crítica, Madrid, 2009.]

———, «The Confusion over Evolution», *The New York Review of Books,* 39 (19): 47-54, 1992.

———, «Life in a Punctuation», *Natural History,* 101 (10): 10-21, 1992.

———, «Confusion over Evolution: An Exchange», *The New York Review of Books,* 40 (1-2): 43-44, 1993.

Graham, George, y Terence Horgan, «Mary Mary Quite Contrary», *Philosophical Studies,* 99 (1): 59-87, 2000.

Greene, Joshua, y Jonathan Cohen, «For the Law, Neuroscience Changes Everything and Nothing», *Philosophical Transactions of the Royal Society,* 359: 1775-1785, 2004.

Harris, Sam, *Free Will,* Free Press, Nueva York, 2012.

Haugeland, John, *Mind Design,* The MIT Press, Cambridge [Estados Unidos], 1981.

———, *Artificial Intelligence: The Very Idea,* The MIT Press, Cambridge [Estados Unidos], 1985. [Versión en español: *La inteligencia artificial,* trad. de Irene Tulli de Firmani, Siglo XXI, México, 1988.]

Hawking, Stephen W., *A Brief History of Time,* Bantam, Nueva York, 1988. [Versión en español: *Historia del tiempo,* trad. de Miguel Ortuño, Crítica,

Barcelona, 1988.]
Hayes, Patrick, «The Naïve Physics Manifesto», en Donald Michie (comp.), *Expert Systems in the Microelectronic Age,* Edinburgh University Press, Edimburgo, 1978.
Heider, Fritz, y Marianne Simmel, «An Experimental Study of Apparent Behavior», *American*
Journal of Psychology, 57 (2): 243-259, 1944.
Hilliard, John Northern, *Card Magic,* Carl W. Jones, Minneapolis, 1938.
Hofstadter, Douglas, *Gödel, Escher, Bach. An Eternal Golden Braid,* Basic Books, Nueva York, 1979. [Versión en español: *Gödel, Escher, Bach. Un eterno y grácil bucle,* Tusquets, Barcelona, 1987.]
———, «Metamagical Themas: Can Inspiration Be Mechanized?», *Scientific American,* [s. n.]: 18-34, 1982. Reimpreso como «On the Seeming Paradox of Mechanizing Creativity», en Douglas Hofstadter, *Metamagical Themas: Questing for the Essence of Mind and Pattern,* Basic Books, Nueva York, 1985, pp. 526-546.
———, *Metamagical Themas: Questing for the Essence of Mind and Pattern,* Basic Books, Nueva York, 1985.
———, *Le Ton beau de Marot: In Praise of the Music of Language,* Basic Books, Nueva York, 1997.
———, *I Am a Strange Loop,* Basic Books, Nueva York, 2007. [Versión en español: *Yo soy un extraño bucle,* trad. de Luis Enrique de Juan, Tusquets, México, 2009.]
———, y Daniel C. Dennett (comps.), *The Mind's I. Fantasies and Reflections on Self and Soul,* Basic Books, Nueva York, 1981.
Holland, John, *Adaptation in Natural and Artificial Systems,* University of Michigan Press, Ann Arbor, 1975.
Hume, David, *A Treatise of Human Nature,* ed. de L. A. Selby-Bigge, Clarendon, Oxford, 1964 [1739]. [Versión en español: *Tratado de la naturaleza humana,* ed. y trad. de Félix Duque, Tecnos, Madrid, 1998.]
Humphrey, Nicholas, «Scientific Shakespeare», *The Guardian,* Londres, 26 agosto de 1987.
———, y Daniel C. Dennett, «Speaking for Ourselves: An Assessment of Multiple Personality Disorder», *Raritan: A Quarterly Review,* 9 (1): 68-98, 1989. Reimpreso con notas a pie de página en *Occasional Paper 8,* Center on Violence and Human Survival, John Jay College of Criminal Justice, The City University of New York, Nueva York, 1991.
Hurley, Matthew, Daniel C. Dennett y Reginald B. Adams Jr., *Inside Jokes: Using*

Humor to Reverse-Engineer the Mind, The MIT Press, Cambridge [Estados Unidos], 2011.

Jackendoff, Ray, *Consciousness and the Computational Mind,* The MIT Press, Cambridge [Estados Unidos], 1987. [Versión en español: *La conciencia y la mente computacional,* trad. de Ana Ardid Gumiel, Visor, Madrid, 1998.]

———, *Patterns in the Mind: Language and Human Nature,* Harvester Wheatsheaf, Londres, 1993, y Basic Books, Nueva York, 1994.

Jackson, Frank, «Epiphenomenal Qualia», *Philosophical Quarterly,* 32 (127): 127-136, 1982.

Kane, Robert, *The Significance of Free Will,* Oxford University Press, Oxford, 1996.

Knobe, Joshua, y Shaun Nichols (comps.), *Experimental Philosophy,* Oxford University Press, Oxford, 2008.

Leibniz, Gottfried Wilhelm F. von, «Monadology», en *The Monadology and Other Philosophical Writings,* trad. al inglés de Robert Latta, Oxford University Press, Oxford, 1898 [1714]. [Versión en español: *Monadología,* trad. de Manuel Fuentes Benot, Aguilar, Buenos Aires, 1957.]

Lenat, Douglas B., y R. V. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the cyc Project,* Addison-Wesley, Reading, 1990.

Lettvin, J. Y., U. Maturana, W. McCulloch y W. Pitts, «What the Frog's Eye Tells the Frog's Brain», *Proceedings of the Institute of Radio Engineers,* 47 (11): 1940-1951, 1959.

Levine, Joseph, «Out of the Closet: A Qualophile Confronts Qualophobia», *Philosophical Topics,* 22 (1-2): 107-126, 1994.

Lewis, David, «Truth in Fiction», *American Philosophical Quarterly,* 15 (1): 37-46, 1978.

Lloyd, M., y H. S. Dybas, «The Periodical Cicada Problem I. Population Ecology», *Evolution,* 20 (2): 132-149, 1966.

Lloyd Morgan, Conwy, *An Introduction to Comparative Psychology,* W. Scott, Londres, 1894.

Ludlow, Peter, Yujin Nagasawa y Daniel Stoljar (comps.), *There's Something about Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument,* The MIT Press, Cambridge [Estados Unidos], 2004.

MacKenzie, Robert B., *The Darwinian Theory of the Transmutation of Species Examined,* Nisbet, Londres, 1868.

Maynard Smith, John, *The Evolution of Sex,* Cambridge University Press, Cambridge [Reino Unido], 1978.

McClelland, Jay, David Rumelhart y The PDP Research Group, *Parallel Distributed Processing : Explorations in the Microstructure of Cognition,* vol. 2, The MIT Press, Cambridge [Estados Unidos], 1986.

Menabrea, Luigi Federico, «Notions sur la machine analytique de M. Charles Babbage», *Bibliothèque Universelle de Genève,* 82 (41): 352-376, 1842. Trad. al inglés de Augusta Ada King, condesa de Lovelace, con notas, en *Scientific Memoirs,* 3: 666-731, 1843.

Meyer, Stephen C., *Signature in the Cell: dna and the Evidence for Intelligent Design,* HarperOne, Nueva York, 2009.

Millikan, Ruth, *Language, Thought and Other Biological Categories,* The MIT Press, Cambridge [Estados Unidos], 1984.

———, *White Queen Psychology and Other Essays for Alice,* The MIT Press, Cambridge [Estados Unidos], 1993.

Nagel, Thomas, «What Is It Like to Be a Bat?», *Philosophical Review,* 83 (4): 435-450, 1974.

———, «Recommendation for Book of the Year», *The Times Literary Supplement,* 27: [s. f.], 2009.

———, «Letter to the Editor», *The Times Literary Supplement,* 1: [s. f.], 2010.

Neugebauer, Otto, «A Babylonian Lunar Ephemeris from Roman Egypt», en Erle Leichty, Maria de J. Ellis y Pamela Gerardi (comps.), *A Scientific Humanist: Studies in Honor of Abraham Sachs,* Occasional Publications of the Samuel Noah Kramer Fund núm. 9, Filadelfia, 1989, pp. 301-304.

Pinker, Steven, *The Blank Slate: The Modern Denial of Human Nature,* Viking Press, Nueva York, 2002. [Versión en español: *La tabla rasa. La negación moderna de la naturaleza humana,* trad. de Roc Filella Escolà, Paidós, Barcelona, 2003.]

Popek, Gerald J., y Robert P. Goldberg, «Formal Requirements for Virtualizable Third Generation Architectures», *Communications of the acm,* 17 (7): 412-421, 1974.

Poundstone, William, *The Recursive Universe: Cosmic Complexity and the Limits of Scientific Knowledge,* William Morrow, Nueva York, 1985.

Putnam, Hilary, «The Meaning of "Meaning"», en Keith Gunderson (comp.), *Language, Mind and Knowledge.* University of Minnesota Press, Minneapolis, 1975, pp. 131-193. Reimpreso en Hilary Putnam, *Mind, Language and Reality,* Cambridge University Press, Cambridge [Estados Unidos], 1975, pp. 215-271.

Quine, W. V. O., *Word and Object,* The MIT Press, Cambridge [Estados Unidos], 1960. [Versión en español: *Palabra y objeto,* trad. de Manuel Sacristán, Herder, Barcelona, 2001.]

———, «Universal Library», en *Quiddities: An Intermittently Philosophical Dictionary,* Harvard University Press, Cambridge [Estados Unidos], 1987.
Rapoport, Anatol, *Fights, Games, and Debates,* University of Michigan Press, Ann Arbor, 1960.
———, «Three Modes of Conflict», *Management Science,* 7 (3): 210-218, 1961.
Ridley, Matt, *The Red Queen: Sex and the Evolution of Human Nature,* Macmillan, Nueva York, 1993.
———, *Nature via Nurture,* Fourth Estate, Londres, 2004. La obra también fue publicada bajo el título de *The Agile Gene: How Nature Turns on Nurture,* HarperCollins, Nueva York, 2003. [Versión en español: *Qué nos hace humanos,* trad. de Teresa Carretero López-Tello e Irene Cifuentes de Castro, Taurus, Madrid, 2004.]
Ross, Amber, *Inconceivable Minds,* tesis de doctorado en Filosofía, University of North Carolina, Chapel Hill, 2013.
Ruina, Andy, «Cornell Ranger, 2011, 4-Legged Bipedal Robot 40.5 Mile Ultra-Marathon Pictures», Cornell University, Ithaca, 2011. Disponible en <http://ruina.tam.cornell.edu>.
Rumelhart, David E., James L. McClelland y The PDP Research Group, *Parallel Distributed Processing : Explorations in the Microstructure of Cognition,* vol. 1, The MIT Press, Cambridge [Estados Unidos], 1986.
Ryder, Dan, Justine Kingsbury y Kenneth Williford (comps.), *Millikan and Her Critics,* Wiley-Blackwell, Oxford, 2013.
Sanford, David H., «Infinity and Vagueness», *Philosophical Review,* 84 (4): 520-535, 1975.
Schönborn, Christoph, «Finding Design in Nature», *The New York Times,* Editorials/op-Ed, Nueva York, 7 de julio de 2005.
Searle, John, «Minds, Brains and Programs», *Behavioral and Brain Sciences,* 3 (3): 417-424, 1980.
———, «The Myth of the Computer», *The New York Review of Books,* 29 (7): 3-6, 1982. (Reseña de *The Mind's I.*)
———, «Turing the Chinese Room», en T. Singh (comp.), *Synthesis of Science and Religion, Critical Essays and Dialogues,* Bhaktivedanta Institute, San Francisco, 1988.
Sellars, Wilfrid, «Philosophy and the Scientific Image of Man», en *Science, Perception and Reality,* Routledge & Kegan Paul, Londres, 1962.
Seung, Sebastian, «The Once and Future Science of Neural Networks», presentado en The Society for Neuroscience Meeting, San Diego, 4 de noviembre de 2007.
Siegel, Lee, *Net of Magic: Wonders and Deceptions in India,* The University of

Chicago Press, Chicago, 1991.
Sims, Karl, «Evolved Virtual Creatures», Karl Sims Website, [s. l.], 1994. Disponible en <http://www.karlsims.com/>.
Strawson, Galen, «Evolution Explains It All for You», *The New York Times,* Archives, Nueva York, 2 de marzo de 2003. (Reseña de Daniel C. Dennett, *Freedom Evolves,* Viking Penguin, Nueva York, 2003.)
———, «Your Move: The Maze of Free Will», *The New York Times,* Opinionator, Nueva York, 22 de julio de 2010.
Thompson, Evan, *Mind in Life,* Belknap Press/ Harvard University Press, Cambridge [Estados Unidos], 2007.
Uexküll, Jakob von, «A Stroll through the Worlds of Animals and Men: A Picture Book of Invisible Worlds», en Claire H. Schiller (comp. y trad.), *Instinctive Behavior: The Development of a Modern Concept,* International Universities Press, Nueva York, 1957 [1934].
Vohs, Kathleen D., y Jonathan W. Schooler, «The Value of Believing in Free Will: Encouraging a Belief in Determinism Increases Cheating», *Psychological Science,* 19 (1): 49-54, 2008.
Von Neumann, John, *Theory of Self-Reproducing Automata,* ed. póstuma por Arthur Burks (comp.), University of Illinois Press, Champaign y Urbana, 1966.
Von Neumann, John, y Oskar Morgenstern, *Theory of Games and Economic Behavior,* Princeton University Press, Princeton, 1944.
Wang, Hao, «A Variant to Turing's Theory of Computing Machines», *Journal of the Association for Computing Machinery,* 4 (1): 63-92, 1957.
Wiggins, David, «Towards a Reasonable Libertarianism», en T. Honderich (comp.), *Essays on Freedom of Action,* Routledge & Kegan Paul, Londres, 1973, pp. 31-63.
Wimsatt, William C., «Randomness and Perceived Randomness in Evolutionary Biology», *Synthese,* 43: 287-290, 1980.
Wolfe, Tom, «Sorry, But Your Soul Just Died», en *Hooking Up,* Farrar, Straus & Giroux, Nueva York, 2000, pp. 89-112.
Wooldridge, Dean, *The Machinery of the Brain,* McGraw-Hill, Nueva York, 1963.
Wright, Robert, *Nonzero: The Logic of Human Destiny,* Pantheon, Nueva York, 2000. [Versión en español: *Nadie pierde. La teoría de juegos y la lógica del destino humano,* trad. de María Luz García de la Hoz, Tusquets, Barcelona, 2005.]
Zahavi, Amotz, «The Theory of Signal Selection and Some of Its Implications», en V. P. Delfino (comp.), *Proceedings of the International Symposium on Biology and Evolution,* Adriatica Editrice, Bari, 1987, pp. 305-327.

Índice analítico

a priori: filosofía como disciplina, 295; verdades, 295-296
Abbott, Edwin A.: 260n
abejas: 173
abogado naval: 34, 34n
abogados: 287
abrelatas: 103 absolutismo: 150
abstracto, abstracción: 143, 176-177, 195, 196, 243, 259-260, 295; autómata celular, 259; centro de gravedad como, 236-237, 238; centro de gravedad narrativa como, 241; generalizaciones, 15
abuelos: 156, 159
aburrimiento: 245
ácaros: 180
accidentes: 18, 161, 181, 195, 279
acción: 130, 174, 177, 242-243, 282, 283, 285; buena información y, 264; elección de, 64, 126, 127, 128, 267, 277
acerca de algo (ser), significado o contenido y: 53, 56, 64n, 118-119, 127
acertijos: 13, 149, 185; de ajedrez, 294; crucigramas, 132-133
ACGT (adenina, citosina, guanina y timina): 129, 156, 157, 196-197
acidez: 54
ácido desoxirribonucleico. *Véase* ADN
ácido universal: 149-150, 200
actitud proposicional: 64n, 116
actitudes: 107n, 116, 287, 299
actividades rutinarias: 221
actos de habla, interpretación de: 242
Adams, Reginald B.: 221, 296n
Adán Y-cromosómico: 181-182
adaptacionismo, adaptaciones: 45-46, 198, 199
Adelardo de Bath: 108
adenina (A): 129, 156, 157, 196-198
ADN: 80, 124, 156-157, 196-198, 233, 269, 284, 301; ácido universal y el, 150n; basura, 26; código para el, 156; descubrimiento del, 259; diferencias y similitudes

en el, 151, 156; errores de copiado en el, 26, 156, 157; estructura del, 33; expresión y replicación del, 259; misterio del, 259; mitocondrial, 180; como receta, 26, 157; de tiburones vaca, 136, 137
África: 180
agentes morales: 276, 278
agentes secretos: 283
agentes: dos cajas negras y, 144; intencionalidad y, 128; libre albedrío y, 252-253, 263-265, 274, 282, 283; morales, 276, 278; postura intencional y, 64n, 66, 69 ; psicología vulgar y, 60, 61-62, 63; racionales, 64, 72, 77; robot gigante y, 127; subpersonales, 69 , 73; en la teoría de juegos, 263, 283
agrandar: 75n
agresividad: 49, 116, 253
agricultura: 178
agua: 59, 64n, 123n, 125, 127, 136, 164; ciclo del, 183, 185
agujeros: 37, 281
aire acondicionado, aparatos de: 104
aislamiento: de las artes y las humanidades, 150; geográfico, 177-178; libre albedrío y, 251; en el mundo de Life, 255, 257; reproductivo, 177
ajedrez: broma de, 296n; computadora que juega al, 63-64, 66-67, 106-107, 108, 127, 191-192, 231, 254, 271-278, 295; verdades *a priori* del, 295-296; verdades de orden superior del, 294 ajetrez, verdades de orden superior del: 294-298
Alain (Émile-Auguste Chartier): 199n alas: 161
Alaska: 268 albatros: 292
albedrío: 72, 274; de las células, 129n alcohólicos, amnesia en: 56
aleatoriedad: la biblioteca de Babel y la, 153, 155; cuántica, 269; dos cajas negras y, 138, 139, 140, 141; en la evolución, 156, 163, 196; al jugar al ajedrez, 192; en Life, 261; en el universo, 254
alegría: 80
alemán: 187-188, 282
alfabeto, orden alfabético: 108, 155, 158, 269 algo y no más bien nada: 43
algoritmos: 22, 108-109, 191, 257, 271; en la evolución, 167-168, 170, 196
alianzas, formar: 127, 128 alineación i: 136
Al-Juarismi: 140
alma: 32, 113, 236, 241; análisis de Platón del, 73 almejas: 68
álter: 239, 240 altruismo: 34n, 46, 235 aluxes: 264
ambiente, medio ambiente: 43-44, 177, 178, 263, 264; ajedrez y, 275; de células, 76; conducta y, 282-287; interno, 58-59; de la rana, 186 ambigüedad, profulidad y: 48

amenazas de muerte, detector de: 54 amenazas de muerte: 54, 55
America (velero): 291
America's Cup: 291 amibas: 157 amígdala: 113, 236
aminoácidos: 189, 261
amnesia, síndrome de Korsakov: 56
Amoeba dubia: 157
amor: enamoramiento, 70, 264; levantamiento de muros y, 150; como palabra, 48-49
analogía: 15; entre genes y palabras, 157-159 análogo: 269, 270
ancestros: 183; comunes, 29, 159, 163, 178, 182; directos en línea materna, 180-181 ancho de banda: 84
andamiajes: 15, 57, 63, 241
anestesiólogos: 33
Anfitrión Diabólico: 18 anhelos: 149
animación, para clases de introducción a la psicología: 62
animales: 126, 183; aprendizaje en los, 115; cerebro de los, 51-52, 54, 77-78; diseño copiado de los, 162; estados como de creencia de los, 69 ; evolución de los, 168-169; genoma de los, 157; imagen manifiesta y los, 57; impredecibilidad de los agentes y los, 263-264; lenguaje de los, 52; mentalidad atribuida a los, 37; psicología vulgar y los, 60, 61, 62; postura intencional y los, 64, 68 , 69; psicología de los, 69 ; significado y los, 51-52; *véanse también animales específicos*
ansiedad: 61, 212, 245
Ant and the Peacock, The (Cronin): 45-46 antenas: 167n
antiesencialismo: 78 antigüedades, compra de: 264 antiintuitivo: 292, 293
antipensamiento, herramienta de: 38; *véase también*
muleta de razonamiento antropocéntrico, sesgo: 54
antropología, antropólogos: 131; auto-, naïf, 292-294; axiomática, 292; naïf sofisticada, 293-294 antropomorfismo: 74, 115, 165
apareamientos interlinaje: 178 apilamiento: 44, 46 apreciación de la filosofía: 299 apreciación: 71, 130
aprendizaje: 80, 82-83, 85, 112, 137, 297; capacidad de, 129; de conceptos en contraste con memorización, 170; conciencia y, 47, 246, 248, 249-250; errores como oportunidad de, 26; evolución y, 26; de la lengua, 80, 82, 112; por partes, 115; psicología vulgar y, 61-62, 68 ; por refuerzo, 27; de robot gigante, 128
aptitud, buena forma: 174, 199, 282
Arabidopsis thaliana: 254 arados: 161
árbol de la vida: 108, 159, 161, 166, 168, 169

árbol de Navidad: 203
árboles: 68 , 83, 135, 159, 160, 161; la conciencia y los, 203; las razones y los, 172
archivos ejecutables: 111 ardillas: 172
argumentos «ancestrales»: 280
argumentos, disyunción en la estructura de los: 176
Aristóteles: 113, 118, 299
aritmética: 102, 114, 172, 173, 218; algoritmos y, 108; división larga, 19, 28, 104, 108, 109; con máquinas de registro, 87, 88, 90-97, 98-99, 100, 231; con máquina de Turing, 100; con máquina de Von Neumann, 100-101; operaciones de coma flotante y, 101, 102
armas: 122; herramientas de pensamiento como, 20, 49
Armstrong, Louis: 213
arn (ácido ribonucleico) mensajero: 197 arqueas: 75, 168, 169
arquitectura: 114, 117; del cerebro, 200; del cerebro de Kaspárov contra Deep Blue, 191; informática, 76; naval, 199; razones para la, 172; de subsunción insectoide, 117n
arquitecturas informáticas: 76 arremangarse: 19
arte, artistas: 41, 150, 192, 196, 291, 299
arte: 150, 290; la ciencia comparada con el, 41-42, 150, 290, 291; hacer que lo conocido resulte extraño y el, 61; pensar afuera de la caja y el, 41; *véanse también artes específicas*
artefactos: 193; ajedrez como, 296; computadora que juega al ajedrez, 127; diseño de, 161-164, 166, 168, 170; humanos como, 130-131; intencionalidad y, 118, 120, 126, 128, 146-147; lenguaje metafórico y, 121, 126, 128; postura intencional y, 64, 65-66, 68 -69; robot gigante como, 126, 129, 130; tragacuartos como, 120, 121; vim como, 214
artífice inteligente, eliminación del: 129 artimañas: 19
«atrapado en el centro de operaciones del robot»: 81-83
ascenso semántico: 19n
ASCII (Código Estándar Estadunidense para
Intercambio de Información), código: 98, 142, 145, 152n, 193n, 197
Asesinato en la catedral (Eliot): 240n
asesinato: 268, 284, 285, 286-287; glutamato y, 32; en la Plaza de Trafalgar, 52-55, 115 asexual, reproducción: 169
asignación de abono, problema de: 29 asociaciones: 83
asteroides: 172, 250, 253, 288
astrocito-neurona, interacciones: 107n
astronomía, astrónomos: 78, 288, 300

atención: 221; efecto del color sobre la, 217-218 atletas innatos: 170-171
átomos: 50, 57, 72, 136, 165, 260, 261; de carbono, 204; de oro, 268-169; de tiburones vaca, 136; división de los, 165 atracadores, creencia en: 53
Austen, Jane: 209, 210
Austin, John: 298 autismo: 63n
autoantropología: axiomática, 292; naïf, filosofía como, 292-294; naïf naïf, 292-293
autocontrol: 183 autofenomenología: 242 autómata celular: 259 automonitoreo de zimbos: 207 autores artificiales: 188-191
autorreplicación: 167, 200, 259-262 autorrepresentación recursiva: 207-208
Auxiliares: 73
ave tejedora: 162
aves: 13, 51, 52n, 59, 263; evolución de las, 159, 160, 161, 162, 171; nidos de las, 161, 162, 171; osos hormigueros comparados con las, 59
avestruces: 292 aviones: 161
avispa *Sphex:* 281, 282 avispas: 59-60, 281, 282
Axelrod, Robert: 34 axiomática, antropología: 292 azar: 156, 160, 188, 267, 270 Babbage, Charles: 96, 113
bacterias: 58, 75, 77, 159, 160, 161, 168, 169, 180; competencias de las, 83; diseño de las, 161, 168; flagelo de las, 169; maomeno, 77; razones de las, 172
Bach, Johann Sebastian: 193, 211
«balada de la pizzería Shakey's, La»: 303 baloncesto: 21
ballenas: 159, 160, 175, 291
«baraja afinada, la»: 223-226
Barrett, Justin: 62 base, de número: 101
bases de datos: 134n; CYC, 117n, 134n, 141n; dos cajas negras, 140-144; sistemas expertos, 140
basset hounds: 182
basura: 195, 298; del ADN, 26; en la biblioteca de
Babel, 161, 166; en el Espacio de Diseño, 161, 166
basura, el noventa por ciento de todo es: 36
Bateson, William: 79, 205, 207, 210, 233, 301
Baum, Eric: 75
Bayes, teorema de: 15
Bayes, Thomas: 15
bayesiano, pensamiento estadístico: 15, 148n bbc: 303
beagles: 182
bebés, primera infancia: 181, 291
Beethoven, Ludwig van: 41, 161

*Behavioral and Brain Sciences (bbs):*70n, 227
Behe, Michael J.: 185n belleza: 294
beneficios, traer: 68
Bennet, Max: 76n
«biblioteca de Babel, La» (Borges): 151-152, 153, 154, 155
biblioteca: de Babel, 151-156, 161, 165, 189; de Mendel, 155-157, 161 bibliotecas: 299
Big Bang Theory, The (serie de televisión): 63n
Big Bang: 153, 267
Big Dog (robot caminante): 164-165 biocombustibles: 200
biología, biólogos: 80, 136, 179, 184, 212, 250, 300, 303; del desarrollo, 64; evolutiva, 64, 70, 162-163, 170, 176; moleculares, 25; «organismos modelo» y la, 256; resistencia a hablar de diseño en, 162-163
biología del desarrollo: 64
biológica, evolución. *Véase* evolución biosfera: 162, 163, 171
bisonte: 178
Blake, William: 128
Blind Watchmaker, The (Dawkins): 156
Block, Ned: 46-47, 213
Bloom, Paul: 251
Bobrow, Daniel: 292n boca: 70, 238
boletos de avión, reservaciones: 107 bombas: 162; corazón como, 130, 131
bombas de intuición: acuñación de la expresión, 16, 205, 227; ingeniería inversa de las, 17-18; introducción a las, 13-23; que quedaron fuera, 303-304; valor de las, 147; *véanse también bombas de intuición específicas*
bondad: 43, 294 bonobús: 29
Borges, Jorge Luis: 151-152, 152n, 154, 155 borrachera: 279
Boston Globe: 294
Brahms, Johannes: 165
Brainstorms (Dennett): 74, 303
Bray, Dennis: 184
brazos: 34, 161; fracturados, yeso para, 105
Brenner, Sidney: 38, 39
brincos de las gacelas: 173-174, 263, 278
Brooks, Rodney: 117n
bucles: 271; de retroalimentación, 263
«Bueno, en ese momento no parecía tan mala idea»: 27
buscar, tratar de conseguir, por robot gigante: 126 búsquedas: 155, 190, 191, 192

Byron, Lord: 96n
caballos: 19n, 128; de la Tierra Gemela, 123-125, 131
«cadena de custodia» de las pruebas en juicios por asesinato: 268
Caenorhabditis elegans: 254 calamar: 254
calcio, canales de: 107n
calculadoras, diferencia entre robots y: 77 cálculo: 15, 204
calentamiento global: 253 calórico: 42
cámaras: 162, 167n caminar: 164 camiones: 68
campanas, hilos para hacer sonar: 113 campanas, sonido de: 113
Campbell, Murray: 191, 193 canguros parlantes: 165
Canis familiaris: 182
Cann, Rebecca: 180 canoas polinesias: 199 canon de parsimonia: 37
Capgras, Jean Marie Joseph: 216
Capgras, mal de: 215, 216-217, 218
«Capitán Autonomía»: 286-287
carácter fenoménico, conocimiento directo del: 220 carbono: 75, 189, 195, 204
carcelero caprichoso: 17-18
Card Magic (Hilliard): 223n
Cargo Cult Lecture (Feynman): 30n caricaturas: 48, 164, 213
carpinteros: 14
cartas, trucos de: 29-30, 31, 223-226, 264 cartesianas, coordenadas: 15
cartesiano, dualismo: 291
Cartesiano, Teatro: 71n, 202 casas: 161, 198
castores, diques de los: 161, 193 catedrales e iglesias: 57, 161 católicos. 32, 33, 163
causal, transmisión (de información genética), 197 causales, relaciones: 179
causalidad, causas: 212, 294; en el ajedrez, 273, 274; aleatoriedad comparada con, 254; dos
cajas negras y, 137-138, 142-143, 146, 148; libre albedrío y, 17, 251-252, 264, 266-267, 282; propiedades semánticas en la, 134; los qualia y la, 213
cd, copiado de: 269-270 ceguera: 61; cortical, 70 células cancerosas: 76 células de levadura: 76, 161
células: 72, 75-76, 160, 189, 235; del cerebro, véase neuronas; ciclos y, 116; copia del genoma en las, 26-27, 157; diseño de, 161, 162, 169, 173; doble hélice, 80, 205; eucariotas, 76, 129n, 168, 169, 180; microtúbulos en las, 33; en el mundo de Life, 255-256, 258, 259, 261; nanorrobots de las proteínas en las, 72, 162; procariotas, 129n
cerebrés: 54, 115-117, 117n cerebro en la cubeta: 71n
cerebro: 30, 112-114, 133, 200, 204, 267, 277, 282, 287, 303; almacenamiento de

creencias y el, 55; almacenamiento de información en el, 115; amputación de partes del, 70-71; de los animales, 51-52, 54, 77-78; atrapado en el centro de operaciones del robot y el, 82-83; comparación de la mente con el, 37, 69 , 113; competencias de los subsistemas del, 69 , 70-71, 72; computadora como metáfora del, 85-86, 113-114, 134n; conciencia y el, 70-71, 202, 203, 204, 220-223, 242, 247, 248; diseño del, 83, 112, 134n; diseño y el, 226; no es»tábula rasa» al nacer, 83; estructura subpersonal del, 117; física vulgar y el, 60-61; funcionalismo homuncular y el, 73-76; ingeniería inversa del, 85, 107, 112; intencionalidad y el, 53-54, 119; de Kaspárov,
191, 192; lenguaje del, 82-83, 115-117; lesión en el, 282; la mente como, 32, 113, 236; como motor semántico y como motor sintáctico, 133-134; necesidades energéticas del, 133; neuronas, 161; nombres gráficos para partes del, 113; operaciones en el, 203; la persona como, 70-71; plasticidad del, 71; como procesador paralelo, 102; procesamiento de información del, 113, 117, 221; psicología vulgar y el, 61-62; de la rana, 185, 186, 187; reconocimiento facial y el, 216; significado y el, 51-56, 62, 69 -72, 109-110, 113, 114, 133-134, 147; sueño y, 184; tejido maravilloso y el, 79-80, 83, 134; trabajo del, 133; trasplante de, 71; tumor en el, 282; el yo y el, 236, 238, 240, 241 cerradura y llave: 113 chaballos: 123-125, 131
Chalmers, David: 220-222, 224, 225
Changeux, Jean-Pierre: 32 chapulines: 173
«Charlie Brown» (personaje de *Peanuts*): 48
Chaucer, Geoffrey: 188 chiflados: 19
chimpancés: 29, 156, 184n
chips de computadora: 103, 104, 110, 112 chistes, bromas: 23, 63n, 71, 221, 236, 252-253
Chittka, Lars: 281, 282
Chomsky, Noam: 31-32, 291-292
Church, Alonzo: 109
Churchland, Patricia: 123
Churchland, Paul: 123 cibernética: 113
ciclos: 183-185; dentro de ciclos, 190
ciencia cognitiva: 31-32, 44, 74, 113, 117, 185, 212, 225, 289; papel del operador maomeno en la, 77 ciencia empírica: 295
ciencia ficción: 36, 128, 129, 141n; robot gigante, 82-83, 125-129 ciencia política: 300
ciencias de la computación, informática, especialistas en: 64, 84, 104, 107n, 230, 231

ciencias sociales: 300
científica, imagen: 57-60, 251, 293-294
científicos, ciencia: 15, 156, 200, 201, 213, 289-292, 299; artes y humanidades comparadas con la, 41-42, 150, 290-291; en chistes y caricaturas, 164; conciencia y, 241-244, 249-250; su desconfianza de las palabras, 20-21; su desdén hacia la filosofía, 22, 135; empírica, 295; errores públicos y, 30; evolución y, 26; experimentos en la, 135-136; experimentos mentales empleados por, 16; filosofía de la, 294, 300; grandes preguntas y la, 22; ignorancia de la, 301; libre albedrío visto por, 251-252, 253, 287-288; magia en el escenario comparada con, 29-30; pensar afuera de la caja y la, 41, 42; psicología vulgar comparada con la, 63; resultados antiintuitivos en la, 293; resultados negativos y la, 30n; sobresimplificación en, 33; supuestos filosóficos de, 25; *véanse también ciencias específicas*
científicos nazis: 284
«cigarra y la hormiga, La», fábula de Esopo: 17 cigarras: 173
cigarros: 53 cinasas: 184
cine, películas: 118, 235, 263, 263n, 284, 303 circuitos neuronales, actividades subpersonales de los: 69 círculo: 206, 247 circunstancias: 64n
cirugía: lobotomía prefrontal, 236; reemplazo de mano o brazo, 34
cisnes: 268
citosina (C): 129, 156, 157, 196-197
«Clapgras, señor», ejemplo imaginario del: 217, 225-226
Clark, A.: 184n clasificador: 59
Clegg, Liam: 265n
coches y manejo de: 61, 104, 165, 237, 279 código: 107n, 198; para ADN, 156; ascii, 98, 141-142, 145, 152n, 193n, 197; compilador y, 108; errores en el, 190n; fuente, *véase* código fuente; máquina, 232; Morse, 146; objeto, 109-110, 111, 112, 145; seudocódigo, 110-111 código de operación: 101
código fuente: 66, 67, 109-112, 118, 272; la habitación china y el, 232, 233; comentarios en, 109-112, 116, 145, 232, 233; sin comentarios, 112; compilación del, 145, 272; sintaxis del, 111 código objeto: 109-110, 111, 112
Cody, Buffalo Bill: 178 cognición: 184, 226-227 cognitivismo: 60
Cohen, Jonathan: 284-286 cohetes espaciales: 161
coletazo: 174 colibríes: 292
coliflor, sabor de la: 211
colisiones: 13, 194, 195, 261, 268-269, 277
color: 58, 251, 252, 287; anomia del, 218; conciencia del, 33, 245-250; daltonismo, 218; ideología vulgar del, 251; imagen manifiesta contra imagen científica y el,

57, 58; preferencias de, 217-218; pruebas de visión del, 219; qualia, 211, 212, 217-220, 225-226, 247-248 color del cielo: 211, 246
coma flotante, operaciones de: 101, 102 comentario crítico, reglas de Rapoport: 34-35 comentarios: 109-112, 116-117, 145, 232, 233 cometas: 46
comida, alimentación: 164, 180, 183; el libre albedrío y la, 283-284; los qualia y la, 211, 218, 219
Common Lisp: 106-107
como si: intencionalidad, 129; propiedades semánticas, 143
compatibilismo: 42, 265, 272-278, 279, 284, 287-288
competencia, competitividad: 49, 72-73, 172, 264, 283; de las neuronas, 75-76; reproducción y, 169-170; de robot gigante, 127, 128 competencias: 80, 108, 231, 234; para el ajedrez, 272, 276; para el color, 220, 225; sin comprensión, 83, 96, 170-174, 192; de emi, 192-193; de la máquina de registro, 99; mentales, suposición tácita de, 56; morales, 276; del motor semántico, 134; el operador maomeno y las, 77, 98; el sphexismo y las, 281, 282; de subsistemas del cerebro, 69 , 70-71, 72; de la unidad de procesamiento, 86-87
compilador: 108, 111-112 complejidad: 185, 233, 259, 263
comportamiento: 130, 162, 253, 259, 263; animal, 37; apropiado, 83; la conciencia y el, 203, 208, 209, 225; de las dos cajas negras, 137-138, 142-143; «fuera de personaje», 238, 239; de imanes, 136; medio ambiente y, 282-287; de razonadores prácticos, predicción del, *véase* postura intencional; razones para el, 172
comprensión: 78, 83, 134, 199, 231; competencia sin, 83, 96, 170-174, 192; de la jerga informática, 84
computadora K de Fujitsu: 102, 103
computadoras: 44, 84-114, 118, 133, 157-158, 269-277; como algoritmos, 108-109; beneficios de las, 169; cuánticas, 102-103; diseño de las, 75, 101-102, 161, 269; elevador manejado por, 59, 104, 109-112, 126; evolución simulada por, 194-196; falta de preocupación de las, 75; funciones cognitivas de la conciencia y las, 221; generadores de números aleatorios en las, 271; invención de las, 73, 77; que juegan al ajedrez, 63-64, 66-67, 106-107, 127, 191-192, 231, 254, 271-278, 295; K de Fujitsu, 102, 103; lenguajes de las, 105-106, 107, 141-142; Mac, 91, 105, 106, 159; máquina de Turing, 13, 85, 99-100, 103, 106, 260; máquina de Von Neumann, 13, 85, 100-101, 260; máquinas de registro, *véase* máquinas de registro; máquinas paralelas, 102, 106, 146; memoria, *véase* memoria de computadora; como metáfora del cerebro, 85-86, 113-114, 134n; el Motor Analítico y las, 96; como motores sintácticos, 143, 144; PC, 91, 105, 106-107; plasticidad de las, 106; portátiles, 13, 85, 86, 99, 100, 104, 106, 231, 260, 269,

270; reducción del tamaño de las, 102; resumen sobre las, 113-114; risc (computadora de conjunto de instrucciones reducidas), 101; siete secretos de las, 96-103; simulaciones de Life por las, 257-258; simulaciones en las, 194-196, 200; tejido maravilloso y las, 79; velocidad de las, 100-103; versatilidad de las, 103; videojuegos para, 194; visión antropomórfica de las, 74; *véase también* máquinas virtuales
computadoras que almacenan programas: 96, 98 cómputos: en las células, 184; al jugar ajedrez, 192; con máquinas de registro, 87-88 comunicación: 242, 289; animal, 52n; cerebro y, 113; explicar de menos y, 40; neuronal, 72; de robot gigante, 128; *véase también* lenguaje; significado o contenido; palabras comunidad: 125
concebir: 301-302; imaginar en contraste con, 205-208 concentración: 217
concestro: 178
conciencia: 22, 46-47, 49, 78, 107n, 114, 148, 150, 176, 251, 287, 302; aprendizaje y la, 47, 246, 248, 249-250; cerebro y, 70-71, 202, 203-204, 221-223, 242, 247, 248; del color, 33, 245-248; la corazonada zombi y la, 202-205; correlatos neurales de la, 33; delfines y, 51; desinflada, 223; las dos cajas negras y la, 144; dos contraimágenes de la, 201-202; los estilos narrativos y la, 210; evolución y, 148, 149-150; flujo de la, 210, 230; función cognitiva de la, 221, 225; herramientas para pensar sobre la, 201-250, 303; la heterofenomenología y la, 241-245; ideología vulgar de la, 251; inventario de los efectos conocidos de la, 225; maomeno, 230; misterio de la, 148, 223, 251, 301; las neuronas y la, 75-76; la persona y la, 73; problema difícil de la, 220-222, 225, 250; problemas «fáciles» de la, 220-221, 224-225, 250; los qualia y la, 211-220, 247-248; resumen sobre la, 249-250; suposiciones sobre la, 42; la *Umwelt* y la, 59; visual, 245-250; el yo y la, 234-241; los zombis filosóficos y la, 203-210
conciencia explicada, La (Dennett): 71n, 223, 304 concupiscible, lo: 73
conducta. *Véase* comportamiento conductismo, conductistas: 27, 60, 291 conductual: 219, 220
Conferencias John Locke: 21 confesar: 128
congelamiento: 37
conjeturas con cierta base: 28 conocido visto como extraño, lo: 61
conocimiento: 25, 69 , 115, 175-176, 177, 209, 246, 247-248, 291, 294; del ajedrez, 295; bases de datos del, 134, 141, 142; diseño y el, 162-163; evolución y el, 26, 150; innato, 17; maomeno, 141, 209; el tragacuartos y el, 120 construcción, grúas para: 167 construcciones booleanas (lógicas): 148n contenido. *Véase* significado o contenido *Content and Consciousness* (Dennett): 73-74

contradicciones: ocultas, 34; obvias, 34; reducción al absurdo, 16, 31-34
contrato social: 17
contratos: 17, 57, 70, 238, 287
control: 49, 265, 283; por el medio, 283
Convención Mundial de Ciencia Ficción (1953): 36
Conway, John Horton: 255-261, 269
Cooper, Sheldon (personaje televisivo): 63n cooperación: 68 , 125, 242, 243, 298; el robot gigante y la, 127
Cope, David: 191, 192-193, 195
copiado: 157; de cd, 269-270; diseño y, 161; errores de, 26, 156, 177; los memes y el, 198-199; en el mundo de Life, 259, 261
copos de nieve: 156, 272 corazón: 85, 130, 131, 162
corazonada zombi: 202-205, 226, 230 correlatos neurales de la conciencia: 33 correo electrónico: 141n, 271
corteza: área v4 en la, 33; arrugada, 113; occipital (visual), 70, 82, 216 corteza arrugada: 113
cosas que se mueven, seguirles la pista conscientemente: 221
Coyne, Jerry: 162 coyotes: 178-179
cpu (unidad de procesamiento central): 139, 231, 273
creación de hechos históricos inertes: 269 creación especial: 187
creacionistas, creacionismo: 38, 39, 130, 170, 185 creatividad: 22, 170, 261; hacer que lo conocido
resulte extraño y la, 6; la inteligencia artificial y la, 189-193; intrusiones espontáneas en la, 194, 195; los memes y la, 199; modelo informático de la, 195-196; pensar afuera de la caja y la, 41-42, 61; solución de problemas y la, 141 creencia, estados como de: 69
creencias: 176, 292-293; adquisición de, 52-53; en el ajedrez, 274; almacenamiento de, 55; compartidas, 52-54, 11; conciencia y, 201, 242-245; definir las, 64-65; dos cajas negras y las, 141, 142, 144, 145; entendimiento y las, 56, 57; falsas, inserción de las, 55-56, 115, 252-253; la heterofenomenología y las, 242-245; implícitas, 116; individuales contra múltiples, 56; intencionalidad y, 118, 119; el libre albedrío y las, 265-265, 282-283, 285; en el mal de Capgras, 216; maomeno, 57, 77, 142, 148, 208; la postura intencional y las, 64, 66, 67-68, 69 , 77, 242-243; la psicología vulgar y las, 60, 130; *de re* y *de dicto*, 303; ser acerca de algo y las, 56, 64n, 119; el significado y las, 51-54, 55-57, 115, 116, 124, 130-131, 137, 141-142, 147 creer: 201
Crick, Francis: 33, 165, 205, 259, 291 crimen: 284-287; el ADN y el, 151
Crimen y castigo (Dostoievski): 209-210 criptografía: 132, 133, 145, 146

cristianismo, el alma en el: 236
criterio: 72
«Cromaphil, doctor»: 218, 219
cromosoma Y: 181-182
cromosomas: 79, 158, 178, 195; Y, 181-182
Cronin, Helena: 45-46 crucigramas: 131-133
cuántica(s): aleatoriedad, 269; física, 150, 229, 233, 268-269; fluctuaciones, 271; generador de números aleatorios cuánticos, 277; informática, 131-132; superposición, 102-103
cuartos de balboa panameño: 120-123, 128, 130 cuartos de dólar estadunidense: 120-122 cuclillos: 171
cuenta de cheques, sacar el saldo de la: 108
cuento del antepasado, El (Dawkins): 158-159, 162
Cuentos de Canterbury (Chaucer): 188 cuerda: 167n
cuerpo humano: 50, 113, 133, 135, 137; el ADN como receta para crear un, 26; descentralización en el, 71; en dispositivo de hibernación, *véase* robot gigante; el yo y el, 238, 239, 241; *véanse también partes específicas del cuerpo*
culpa, asignación de: 29, 279
cultura: 150, 293; de escritura científica, 20; los memes y la, 198, 200
«curare-cum-amnestic»: 304 curiosidad: 51, 301
Curiosity (explorador robótico): 167n CYC: 117n, 134n, 141n
Dahlbom, Bo: 13
daño cerebral, mal de Capgras y: 216
Darwin, Charles: 45, 77, 108, 149, 150, 160, 163, 168, 170, 171, 186, 205, 231, 261, 280, 291
«Darwin's Strange Inversion of Reasoning» (Dennett): 20
darwinianos, espacios: 303
darwinismo: 191, 200; ciclo en el, 183-185
David Copperfield (Dickens): 155, 158
Davidson, Donald: 135, 137
Dawkins, Richard: 49, 128, 129, 156, 158-159, 162, 178, 193, 198
De Vries, Peter: 154 decágono: 206
Deep Blue: 191-192
Dehaene, S.: 214n delfines: 37, 51, 52n, 124
Demócrito: 266
Dennett, Daniel C. (autor): 16, 17, 20n, 35, 35n, 45-46, 47, 62-63, 64-65, 70n, 71, 73-74, 80n, 108, 117n, 141n, 143n, 185, 186, 190n, 198, 201-202, 207, 211, 212-213, 222, 223, 227-228, 231, 243, 246, 247-248, 260n, 263-264, 285,

296n, 303-304; estilo de escritura de, 20, 21; John Locke Lectures, 21
Denton, Michael: 152n Denys, Damiaan: 252 depredadores: 173, 174 depuración: 111, 112
desarrollo: 26, 69 , 76, 83, 112, 129, 137, 195 desastres, previsión de: 277-278
Descartes, René: 113, 201, 229n, 291, 299; concebir en contraste con imaginar y, 205-206; coordenadas cartesianas de, 15; genio maligno de, 16-17, 212
descendiente de coyote y perro: 178-179 descendiente de lobo y perro: 178-179 descentralización: en el cuerpo, 71; de internet, 71 descompilar: 145
desconfianza: 245
descubrimiento: 184, 195, 252, 253, 259, 261, 283, 290, 292; en el ajedrez, 275, 276; antiintuitivo, 293 deseo: 62, 80
deseos: 51, 60, 116, 190, 285; definición de, 64-65; por herencia genética, 59-60; la heterofenomenología y los, 242-243; la intencionalidad y los, 118-119; la postura intencional y los, 64, 66, 68 , 69 , 77, 242-243; el robot gigante y los, 126; ser acerca de algo y los, 64n, 118-119 despertadores: 65, 66 destino: 267
DeSylva, Buddy: 185
Detección de Bolitas Primaria: 186 detección en robot gigante: 126 detector de billetes falsos: 133-134
determinismo: 140, 186, 254-262, 269; el ajedrez por computadora y el, 271-278; digital, 269; la elección y el, 271-278; el libre albedrío y el, 42, 254-261, 265-267, 271-278, 279, 284, 285, 287-288; Life y el, 254-261
deus ex machina: 167
Dewdney, A. K.: 260n Dewey John: 36 diablillos: 62
diagramas de flujo: 88-91, 93-96, 97 diagramas: 118
Diario de viaje de un naturalista alrededor del mundo en el S. M. Beagle (Darwin): 291
dicotomías, falsas: 43-44 dictado, tomar: 107
dientes, oro en los: 267-268 dientes para fuera, de: 19, 60
Diesel, ciclo de: 183
Diesel, Rudolf: 183 digitalizado: 260, 269-270
Dilbert: 44
dilema del tranvía: 254
dinero: 247; valor del, 213-214; *véase también* dólares; tipo de cambio, paridad; cuartos de balboa panameño; cuartos de dólar estadunidense
dinero de verdad: 213-215 dinosaurios: 46, 159, 292
Dios: 118, 151, 190, 191, 264, 265, 277, 281; la bondad y, 43; existencia de, 37; misteriosos caminos de, 301; la navaja de Ockham y, 37
dioses: 167, 264 diques: 161, 193

discernimiento: 77, 220; del color, 218, 220, 225; en el robot gigante, 126, 127; la selección natural y el, 130
Discurso del método (Descartes): 229n
Diseñador Inteligente: 114, 130, 148, 163, 165, 166, 200, 301
diseño inteligente: 130, 160, 163, 165, 169, 185, 199
diseño: 26, 27, 128, 135, 160-170, 195-196, 200, 285-286; algoritmos y, 109; biológico, 128-129; centro de gravedad en, 237; del cerebro, 83, 112, 134n; de computadoras, 75, 102, 161, 269; cultura y, 199; de CYC, 117n; evolución virtual y, 194-195; excelencia o racionalidad del, 67; genes como beneficiarios del, 129; hechos históricos inertes y, 269; de los humanos, 129-130; I+D: 160-162, 166, 168; inteligente, 130, 160, 162, 163, 164-165, 169, 185, 199; jugar al ajedrez y, 191; en Life, 258-259, 261; malo, 165; maomeno, 77, 194-195, 261; de máquinas, 103; naturaleza de las restricciones del, 133; novedad en, 161; que nunca evolucionó y nunca se construyó, 165; de programas de ajedrez de computadora, 274-275, 276; real en contraste con aparente, 162-163, 164, 168; responsabilidad por el, 279; de robot gigante, 126, 127, 128; selección natural y, 29, 83, 129-130, 160-170, 171, 172; de los tragacuartos, 119-120, 121 diseñoide, uso del término: 162, 164 disminución, registro de: 87
disposiciones: 71, 220 disposiciones (instrucciones): 104 distracción: 80
disyunción en la estructura de argumentos: 176-177 diversidad: 167
división larga: 19, 28, 104, 108; la neutralidad del sustrato y la, 109
doble hélice: 80, 205, 207 dólares virtuales: 58
dólares: 37, 58, 213-214, 251; auténticos y falsos, 133-134
dolor: 34, 70, 245
Don Giovanni (Mozart): 188
«¿Dónde estoy?»: 303
dos cajas negras: 137-148, 173-174, 198
2001: Odisea del espacio (película): 71
Dostoievski, Fiódor: 209-210
Doyle, Arthur Conan: 237-238
Dreben, Burton: 298 drogas: 18
dualistas, dualismo: 32, 246, 251, 291 dudar: 128
duendes: 62, 236, 264
Dulces sueños (Dennett): 304
Eccles, sir John: 32-33
«echar un volado», volados: 267, 272, 275, 276 ecolocalización en murciélagos: 160, 162 economía: 287, 300; de la neuroquímica, 76; de robot gigante, 126, 127; el vim y la, 214 economistas: 115, 214

Eddington, sir Arthur: 57 efemérides: 187
eficiencia: 27, 164, 165, 169, 189, 192, 194, 199, 272 ego: 73, 241
egoísmo: 235
egoístas, genes: 128-129, 148 egoístas, neuronas: 76
Eiffel, Gustav: 188
Eiffel, torre: 188
Eigen, Manfred: 152n
Einstein, Albert: 16, 150-151, 251
élan vital: 213, 214
elección: 17, 271-278, 283, 284; de acción, 64, 126,
127, 128, 267, 277-278; el determinismo y la, 271-278; la selección natural y la, 130, 169, 199; el trato faustiano y la, 290-292 electroencefalogramas, registros de los: 184, 244 electrones: 19n, 42, 57, 67, 152, 243, 269 elefantes: 160, 182, 282
elevadoristas: 110, 112 elevadores, fabricantes de: 110
elevadores manejados por computadora: 59, 104, 110-112, 126
Eliot, T. S.: 240n Elwes, Cary: 263n Ellis, Haydn: 217 ello: 73 embriologías: 158
emi (experimentos sobre Inteligencia Musical): 191, 192-193, 195
emociones: 50, 288; el color y las, 217-218, 225-226, 247; cometer errores y las, 27-28; dinámicas de procesos oponentes de las, 76; el reconocimiento facial y las, 216
enciclopedias: 117n, 144 encriptamiento: 145, 146
energía: 75, 133, 160, 180, 181; ciclos y la, 183, 184; eficiencia de la, 164, 189; para el robot gigante, 125-126, 127 enfermedad: 181, 198 enojo: 27-28, 63 ensamblajes aleatorios: 194 ensayo y error: 26, 30, 160 ensayos ciegos: 26
ensayos previsores: 26 entendedor: 114
entendimiento, entender: 107n, 171, 274, 289; el cerebro y el, 71, 85, 113-114; la conciencia y el, 201, 243; las creencias y el, 56, 57; la evolución y el, 26; a los grandes filósofos, 299; la inteligencia artificial y el, 233-234; maomeno, 77-78, 117, 141, 174, 229; nuevas oraciones, 68; de perros, 51; la prueba de Turing y el, 229; de la responsabilidad moral, 288; *véase también* comprensión
enzimas: 72, 156
enzimas correctoras en el ADN: 72 epidemiología: 200 epifenomenalismo: 143n epigramas: 41
epistemología: 177, 200, 207n
equilibrio puntuado: 43
Erler, Albert: 121 erramientos: 190
errores: 147, 270; en copiado de ADN, 26, 156, 177; etiquetas y, 15; tipográficos,

153; de tragacuartos, 121, 122
errores, cometer: 24-31, 60, 148, 150, 211, 248; buenos, 26-31; como clave para progresar, 25; correcciones y, 31, 249; cultivar el hábito de, 25; filósofos como especialistas en, 25-26; sobre el libre albedrío, 251, 252; como oportunidades, 26, 27-28, 30; repasar el recuerdo de, 27-28; del robot gigante, 126; el tragacuartos y, 119-120; vergüenza y enojo como reacción a, 27-28
escaleras, andamiaje comparado con: 15 escanear: 98, 235
escritura científica, cultura de la: 20 esencia: 136, 176, 179
esencialismo: 176, 187 eslabones perdidos: 184n
Esopo: 16n; fábulas de, 16, 17, 19-20
Espacio de Diseño: 43, 161, 162, 165-170, 187, 190, 191
espacio lógico: 148, 155
espacio: 79, 238, 252, 303; de Diseño, 43, 161, 162, 165-170, 187, 190; hiper-, 155; en el mundo de Life, 257, 259-260
español: 71, 82, 107, 118, 152n
especiación: 177-182, 189; alopátrica, 177-178 especies: 177, 181, 212, 254; ADN de las, 26, 151, 156; como maomeno concepto, 178; organismos individuales comparados con las, 180; origen de las, 46; reproducción sexual de las, 169; la teoría de juegos y las, 263
especificaciones: 107n, 117, 126, 259-260 esperanzas: 149, 241, 288
esperma: 27, 182, 196, 197, 198 espíritu: 241
«espíritus animales» (líquido cefalorraquídeo): 113n esponjas: 172
esquimales: 162, 177
¿Está usted de broma, señor Feynman? (Feynman): 14, 24
estadística, métodos estadísticos: 15, 141n, 148n, 273, 295 estado c: 119-123 estafadores: 240
estilos narrativos: 209-210
estrella de mar: 59, 77 estrés, psicología vulgar: 62 estructura lógica: 109, 117 estructuras nerviosas: 52
estudiantes de licenciatura: 21, 84, 245, 297, 298; como público señuelo, 40-41 estudiantes de posgrado: 21, 297, 298
Estudio de Software Curricular: 86n, 93
éter: 42
ética, expertos en: 34, 150, 254, 300; meta-, 19 etiquetas, rótulos: 15, 81, 82, 83, 116-117; de
advertencia, 15, 23, 46; palabras como, 60; el robot gigante y las, 82-83
eucariontes o células eucariotas: 75-76, 129n, 168-169, 180 euros: 213-214
Eva: 265; Mitocondrial, 179-181 evaluación en ajedrez: 272-273

Evanescente, Evanescentemente: 71n, 153, 157, 166, 189
Everest, monte: 165
evolución: 22, 50, 114, 137, 265, 291, 292, 302, 303; algoritmos en la, 168, 170, 195-196; del altruismo, 34n, 46; apilamiento y la, 44; ciclos en la, 183-185; la conciencia y la, 148, 150; cultural, 198-200; el equilibrio puntuado y la, 43; errores invisibles de la, 30; especiación en la, 177-182; folclórica, 167; geológicamente instantánea, 43-44; gradualismo en la, 43-44, 77, 175-176, 186, 261; herramientas para pensar sobre la, 149-200, 303; historia de la, 133, 146n, 180; de la intencionalidad de los genes, 129; el Mamífero Primario y la, 175-177, 179, 186; el masbienismo y la, 43; los memes y la, 198-199; las mutaciones y la, 26, 261; el Pasodoble de Gould y la, 45-46; la postura intencional y la, 69 -70; punto de vista de Gould sobre la, 43-46; resumen sobre la, 200; significado relacionado con la, 26, 130, 147, 186, 187; simulaciones por computadora de la, 194-196; la teoría de juegos y la, 263; la teoría de las señales costosas en la, 174; teorías de Lamarck sobre la, 291; *véase también* selección natural
evolución cultural: 198-200
«Evolution of Reasons, The» (Dennett): 130n, 185n evolutiva, biología: 64, 70, 162-163, 170, 176
Evolved Virtual Creatures (Sims): 194-195 exaptación: 186
excelencia: 167
expectativas: 64, 116, 238
experiencia: 52, 83, 115, 122, 143-144, 193; conciencia y, 201, 221, 236, 243, 244-246, 248; los qualia de la, 212, 213, 220; subjetiva, 242; el yo y la, 239, 241
experimentación: 161, 241-245, 247, 297 experimentos mentales: centro de operaciones del robot, 81; cerebro en la cubeta, 71n; científicos, 16; definición de los, 16; que imitan experimentos científicos, 135; reducción al absurdo, 16; del trasplante de cerebro, 71; utilidad de los, 137; *véanse también bombas de intuición y temas específicos*
explicación: en el ajedrez, 273, 274, 276; explicar de menos, 40; fáciles de entender, 21, 40-41
explicar de menos: 40 exponente de un número: 101
«expresiones intencionales»: 64n extinción masiva: 45
extinción: 27, 45, 159, 161, 178, 182 extrapolacionismo: 45
Extremada, Extremadamente: 153-155, 156, 157, 165, 188, 189, 191
Fabre, Jean-Henri: 281 fama: 202
Fanshawe, J. H.: 174n fantasía: 241 fantasmas: 62
fantasmas en la máquina: 113 faustiano, trato: 290-292
Fechner, Theodor: 152n felicitaciones de cumpleaños: 54 fenomenología: 242

Fenomenología (con «F» mayúscula): 242 fenotipo: 193
ferrocarril transcontinental: 178
Feynman, Richard: 14, 24, 30n, 254 fibras nerviosas: 113
ficción de teórico: 237-238, 243 filosofía del lenguaje ordinario: 298 filosofía experimental: 293 filosóficos, zombis: 203-210, 222
filósofo(s), filosofía: como autoantropología naïf, 292-294; de la ciencia, 294, 300; definición de Sellars de la, 58; desdén de los científicos hacia la, 22, 135; como disciplina *a priori*, 294-295; su enseñanza a estudiantes de licenciatura y de posgrado, comparación, 21; cómo es ser un, 289-300; como especialistas en errores, 25-26; evolución y la, 26; experimental, 293; experimentos mentales empleados por, 16; hacer que lo conocido resulte extraño y los, 61; herramientas de pensamiento hechas por, 15; historia de la, 25, 299-300; lenguaje de los, 18-21; la Ley de Sturgeon y la, 36; lo mejor de la, 299-300; mejores contribuciones de los, 22; el papel de las bombas de intuición en la, 16-17; pensar afuera de la caja y la, 41, 42; ponerse meta- y la, 18; reflexión sobre ser, 23; trato
faustiano y la, 290-292; *véanse también personas y temas específicos*
«Finding Design in Nature» (Schönborn): 163 firmas legales: 235
firmas, escaneo de: 235
*firmware:*98, 112
Fischer, Bobby: 192
física, físicos: 58, 86, 134, 149, 172, 250, 300; bidimensional, 259-260; cuántica, 150, 229, 233, 268-269; de Life, 255, 257, 259, 261; leyes de la, 65, 66, 120, 121-122, 133, 137, 287; naïf (o vulgar), 60-61, 292, 294; preguntas del tipo del Hombre del pantano y la, 36; soluciones laboriosas en contraste con las fáciles y la, 13; trabajo de Descartes sobre la, 291; virtual, 194
fisiológicamente posible: 218
Fitch, Tecumseh: 129n FitzGibbon, C. D.: 174n *Flatland* (Abbott): 260n flogisto: 42
flujo de conciencia: 210, 230 flujo de inconciencia: 230
Fodor, Jerry: 31, 46n, 52n, 56n, 82n, 119, 120, 121, 123, 130n, 186, 265
«For the Law, Neuroscience Changes Everything and Nothing» (Greene y Cohen): 284
formas, razones y: 172, 173 formato midi: 193
fosfato: 184 fósiles: 204
fotografías: 53; las máquinas de registro y las, 98 fotones: 50, 204, 261, 283
fractura de brazo, yeso para: 105 francés: 107, 282
Frankenstein, doctor: 188-191
«Frankfurt, ejemplos de»: 283

Frankfurt, Harry: 283
Free Will (Harris): 282
Freedom Evolves (Dennett): 35
Freud, Sigmund: 60, 73, 291
Frith, Chris: 217, 251
fronteras: 176, 193; el amor y las, 150; de los organismos, 180
Fuente de la Eterna Juventud: 64 fuerte, inteligencia artificial: 226-234 fuerza de voluntad: 14, 283-284 fuerza psicológica: 264
fumar: 53
función: 131, 148, 186, 187, 195-196; semejanzas de la, 162
función cognitiva: 77, 221, 225 funcional, análisis: 213 funcional, eficiencia: 199 funcional, interpretación: 131 funcional, poder, del dólar: 214
funcional, resonancia magnética: 222, 244 funcionales, diferencias: 210 funcionalidad: 171
funcionalismo homuncular: 74-76, 107, 129n futuro: 133, 135, 277, 288, 289; el Hombre del Pantano y el, 135, 137
gacelas: 174, 263, 278
galaxias: 149, 189; en la biblioteca de Babel, 154-155, 156, 189; de genomas, 156; de pixeles, 260, 261
Galileo: 137 galleta: 60
ganar: 266-267; en el ajedrez, 272-275, 295 ganchos celestes: 166-168, 170
Gardner, Martin: 296n gato, juego de: 108 gatos: 37, 198
Gaudí, Antoni: 172
Geller, Uri: 78, 79, 80
Gell-Man, Murray: 228-229 gemelos idénticos: 156
generadores de números aleatorios: 270-271; cuánticos, 276; generadores de números seudoaleatorios, 271, 273, 275, 277
generalizaciones: 15, 174, 179, 259
genes: 156-159, 161, 182n, 196-199, 205; como beneficiarios del diseño, 129; egoístas, 129, 148; intencionalidad de los, 128-129, 148; como palabras o como subrutinas, 157-159; punto de vista de Bateson sobre los, 79-80, 205, 207; en la reproducción sexual, 169
genética, ingeniería: 170
genética: 59, 79, 80, 182; la imagen manifiesta y la, 59-60; molecular, 200; la psicología vulgar y la, 61-62; replicación, 185; *véase también* ADN genio: 14, 61, 150, 166, 191, 199, 200
genios: 17, 62, 136
genomas: 27, 151, 155-159, 196-197, 254, 261-262; evolución virtual y los, 194-

195 geografía: 177, 295
geología: 300 geometría: 16, 155, 173 geranios: 182
Gerente del Hospital: 18 gérmenes: 42
Gershwin, George: 42, 185, 289 gestos: 63n
Gibson, J. J.: 59n gigahertz: 84
Gilbert, William: 242
girar las perillas: 18, 54, 123n, 147, 148, 190-191, 283; la habitación china y el, 227, 233-234; Mary, la científica del color y el, 246-248; los niños del Brasil y el, 284-287
glaciares: 268
globos oculares, prótesis: 70 glutamato: 32
glutamato monosódico: 219 gnomos: 62
Gödel, Kurt: 86, 108
Gödel, teorema de: 166
Godfrey-Smith, Peter: 196-197, 303, 304
gofai (inteligencia artificial a la antigua): 74, 148n
Goffman, Erving: 239
Goldberg, Robert: 104
Google: 61, 303
Gore, Al: 216
Gould, Glenn: 211
Gould, Pasodoble de: 45-46
Gould, Stephen Jay: 43-46, 173; el apilamiento y, 44, 46; el masbienismo y, 43-44, 46 GPS, aparato de: 28, 29
Graham, George: 220, 248
gramática: 60, 70, 82, 153; reglas de la, 112, 199 gran danés: 182
Grandin, Temple: 63n
gravedad narrativa, el yo como centro de: 236-241 gravedad, centros de: 172, 237-239
Grecia antigua: dramaturgia de la, 167; filosofía de, 16, 36, 38
Greene, Joshua: 284-287 grillos: 281
grúas: 167-170; cascadas de, 167, 193; definición de, 170; la ingeniería genética como, 170; la sexualidad como, 169-170, 200; la simbiosis como, 168, 200
grúas celestes: 167n
guanina (G): 129, 156, 157, 196-197
guardián entre el centeno, *El* (Salinger): 158, 209
Guardianes: 73
guerra de las galaxias, La (película): 203

Guha, R. V.: 117n, 141 guiar: 85
guijarros: 165
Guillermo de Ockham (u Occam): 37 gusanos: 59, 83, 159, 180, 182, 254, 282 gusto: 57, 70, 211, 219
habilidades: 274
habitación china, la: 16, 205, 226-234 hábitat: 282
hábitos: 25, 239
hablar: 18, 59, 107n, 221, 233
Hacedor de Viudas (revólver): 179
Hacker, P. M. S.: 76n hackers: 145, 190n hachas: 66, 161 hadas: 62, 264
Haggard, Patrick: 251
Hal (computadora inteligente): 71
Hameroff, Stuart: 33-34
Hamilton, William: 34n
Hamlet (Shakespeare): 196, 290
Hao Wang: 86, 100
hardware: del cerebro, 21n, 112; de computadora, 75, 84, 101, 104-107, 231; de máquina real, 104-105
Harris, Sam: 282
Haugeland, John: 74n, 134n
Hawking, Stephen: 152n, 251
Hayes, Patrick: 292-294
Hebb, Donald: 296, 300
hechos: 289, 291; conductuales o disposicionales, 133; las dos cajas negras y los, 142, 146; empíricos, 295; históricos inertes, 261, 267-271; inconvenientes, ocultamiento de los, 38-39, 41; más profundos, 131, 133, 147, 150; triviales, 133; uno solo, 115
Heider, Fritz: 62
«hermano mayor que vive en Cleveland»: 55-56 herramientas: de artesanos, 14; de pensamiento, *véase* pensamiento, herramientas del herreros: 14-15
heterofenomenología: 219, 241-245 hexágono: 173
hibridación: 179 hielo: 59, 136
Higgs, bosones de: 50
Hilliard, John Northern: 223-224
Hillis, Danny: 146 hiperastronómico: 153n hiperespacio: 155 hipocampo: 113, 135, 296 hipótesis: 65
historia política: 299
historia: 178, 182n, 295; evolutiva, 133, 146n, 180; de la filosofía, 25, 299-300; el

Hombre del Pantano y la, 135, 137; personal, 133, 179; significado de términos y la, 124; de la vida, 146-147
Hitler, Adolf, clones de: 284
Hobbes, Thomas: 17, 154n
Hofstadter, Douglas: 17, 54, 123, 194, 227; herramientas, su lista de, 19-20, 60; pensar afuera de la caja y, 20, 41; sphexismo, su acuñación de, 20, 281
holismo: 132; de lo mental (o de lo intencional), 56n
Holland, John: 169n
Holmes, Sherlock: 38, 237
Hombre del Pantano conoce a un tiburón vaca: 134-137
Homero: 16n
Homo sapiens: 58, 156, 177, 181
«Homunculi Rule» (Dennett): 304 homunculus: 73-76, 82-83, 107, 112 hongos: 76, 180
Honorable Compañía de Tintoreros: 268n Honorable Compañía de Viticultores: 268n Horgan, Terry: 220-248
hormigas: 75; como término masa, 59 hotel, real y virtual: 194
huellas digitales: 156, 269
huesos, células de los: 157
huésped (yo dominante): 239, 240
huevos: 37; de aves, 171; de avispa, 281; óvulos humanos, 27, 196, 197, 198
Hull, Ralph: 222-225 humanidades: 150
humanos: 29, 44, 189, 293; como artefactos, 130-131; comparados con el ratón, 158; diseño de, 129-130; como máquinas de supervivencia, 128-129
Hume, David: 236, 245 humor: 150, 221
Humphrey, Nicholas: 188, 239, 240, 290-293 huracanes, simuladores de: 128
Hurley, Matthew: 221n, 296n
Husserl, Edmund: 242
I+D (investigación y desarrollo): 160-162, 166, 168, 188, 189-192
ia. *Véase* inteligencia artificial
ibm: Deep Blue, 191; Watson, 54-55, 141n, 144 icoságono: 206
ideas: 15, 38, 71n, 116, 299, 301; la evolución y las, 26; malas, poner en evidencia las, 148 idénticos, gemelos: 156
identificación equivocada, el tragacuartos y la: 119-120, 121
Ignorancia Absoluta: 160, 166, 170-171
ilusión: 58, 80, 107n, 212; de contraste de color, 213; libre albedrío como, 251-253; visual, 107n, 249 imagen: 116; científica, 57-60, 251, 293-294; manifiesta, *véase* imagen manifiesta

imagen manifiesta: 57-60, 251, 287, 292n, 293, 294; metafísica de la, 294; motivo más importante en la, *véase* vulgar, psicología
imaginación: 21, 22, 69 , 70, 136, 234; la biblioteca de Babel y la, 152; la conciencia y la, 201, 202, 203, 210, 226, 247-248; de Descartes, 113n; las dimensiones y la, 161; falta de, 38; la genética y la, 80; inhibición de la, 136; las máquinas virtuales y la, 104; los memes y la, 199
imaginar en contraste con concebir: 205-208 imanes, magnetismo: 136, 242
imitación: del cerebro, 134; en experimentos mentales, 135
imparcialidad de loterías: 266-267 impostores, el mal de Capgras y los: 216 impredecibilidad práctica: 265
impredecibilidad: 254, 262-265; absoluta, 265; práctica, 265 imprenta, presas de: 161 incompatibilismo: 35, 265
incremento, registro de (Inc): 87, 89, 90, 91, 96, 100, 101, 113; versión neuronal del, 102 indeterminación, indeterminismo: 133, 150-151, 186, 237, 265, 267, 270, 276-279, 284 indexación finst: 221
indicadores: 118, 145 individualidad: 290
inertes, hechos históricos: 261, 267-271 infección: 180
inferencia, motores de: 117n, 134n, 140, 141 infertilidad: 178, 197-198
infinito: 271; la biblioteca de Babel y el, 152; de los mamíferos, 175
información: 14, 86, 124, 126, 189, 196, 263; almacenamiento de la, en el cerebro, 115; buena, 264-265; ciclos de la, 184; la conciencia y la, 202; demasiada, 67, 81; dólares hechos de, 58; física, 245-248; genética, 80, 195, 197, 198; para jugar ajedrez, 127; las máquinas virtuales y la, 104; los memes y la, 199; oportuna, 17; perceptual, 221; procesada por computadoras, 103, 104; procesamiento de la, en el cerebro, 113, 117, 221; la psicología vulgar y la, 60; sensorial, 54, 72; sistemas de, 67; usada por zombis, 207; visual, 102; el yo y la, 235
ingeniería: 107n, 167n, 200, 213, 300; bidimensional, 260; del detector de amenazas de muerte, 54; genética, 170; inversa, *véase* ingeniería inversa; prueba de concepto en, 75n; el robot gigante y la, 125, 126, 128, 129; el tragacuartos y la, 120; la *Umwelt* y la, 59
ingeniería inversa: 163; de las bombas de intuición, 17-18; del cerebro, 85, 107, 112; del código objeto, 145; de las dos cajas negras, 145; de la habitación china, 227
ingenieros: 69 , 128, 129, 133-134; lenguaje de los, 126; para programas de computadora, 67; *software,*69
Inglaterra, reina de: 268
inglés: 74, 146, 158, 245; la biblioteca de Babel y el, 151, 152, 153; la habitación

china y el, 231-233; del Hombre del Pantano, 135, 136; como lengua internacional de la ciencia, 20; Máquina Virtual del (MVI), 112; reglas gramaticales del, 112
injusticia «cósmica»: 277
innato: conocimiento, 17; postura intencional, 67-68; psicología vulgar, 61-62, 67-68 insectoide, arquitecturas de subsunción: 117n insectos: 37, 59, 274, 282
instrucción disminuye o ramifica (dir): 87-90, 91, 96, 97, 101, 113; versión neuronal de la, 102 instrucción finaliza: 86-87, 96
inteligencia: 27, 129, 231, 291, 296-297; artificial, *véase* inteligencia artificial; el cerebro y la, 113-114; maomeno, 77; la prueba de Turing y la, 229 inteligencia artificial: 29, 54-55, 73-75, 107n, 123, 128, 292n; a la antigua, 73-74, 148n; la creatividad y la, 188-193; CYC, 117n, 134n; fuerte, 226-234; primer principio, 134n; problemas de juguete en la, 110, 254; Watson, 54, 55, 141n, 144; *véase también* robots intencionalidad: 53-55, 64n, 118-129, 303; *como si,* 129; derivada, 118-123, 126, 128, 131, 147; de
los genes, 129, 148; maomeno, 126, 148; original (o intrínseca), 118-119, 121, 123-125, 128, 129, 131, 146-147, 148; el tragacuartos y la, 121, 123 intencionalidad derivada: 118-123, 126, 128, 131, 147
«Intentional Systems in Cognitive Ethology» (Dennett): 70n
«Intentional Systems Theory» (Dennett): 70n interesante(s): artículo sobre física naïf, 292; factores, del ajedrez, 272-273, 276; libros, en la biblioteca de Babel, 152; problemas, 141; reglas a priori del ajedrez, 295
intereses, de los humanos contra los de los genes: 129
intermitente (semáforo): 256, 257-258
internet: descentralización de, 71; Java e, 105-106 interpretación: 242, 243, 244; funcional, 131; intencional, 133; mejor, 147
intrínseco(a): intencionalidad, *véase* original (o intrínseca), intencionalidad; propiedades fenoménicas, 214; qualia, 211-212, 213, 226; valor económico, 214
intromisiones espontáneas: 194, 195
intuición: 192, 202, 293; conciencia e, 222, 225; como variable, 135-136 invento: 183, 184
inversión: 42
investigación de mercado: 266 invisibilidad: 30, 38, 60, 264
irascible, lo: 73
Ishihara, test de las cartas de: 217 isla de Cornwallis: 177
Israel, David: 186
«J'ai faim/Tengo hambre»: 53
Jackendoff, Ray: 42, 112n, 201n

Jackson, Frank: 245-248
Jacob, François: 76
James, William: 24
Japón: 102, 164 jazz: 41, 213
Jeffries, Lynn: 303 jibias: 51
jirafas: 146n
Johnson, Lyndon: 52n jonrón: 60
Joplin, Scott: 193
joyas en el bote de la basura: 17 jueces: 287
jugada forzada: 67 jugada profunda: 275
Julio César: 267-268 justicia: 19n, 294
Kane, Robert: 35
Kant, Immanuel: 288, 299
Kaspárov, Gary: 191-193
Kennedy, John F.: 153
Kepler, leyes de: 37
King's Gambit: 296n
Kinsbourne, Marcel: 13
Klee, Paul: 61n Knobe, J.: 293n Koch, Christof: 33
Korsakov, síndrome de: 56
Krebs, ciclo de: 183
Krebs, Hans: 183
Kripke, Saul: 119, 120, 121, 123
La Jolla, reuniones de té: 33 ladrillo, aventar un: 277, 278
Lamarck, sus teorías sobre la evolución: 291 lámparas de pie: 237
Language of Thought, The (Fodor): 52n
Laplace, Pierre: 257
lectura de pruebas tipográficas: 190
lectura: de código por la computadora, 103; maomeno, 98; por la máquina de Turing, 100; los programas OCR (de Reconocimiento Óptico de Caracteres) y la, 98
Leibniz, Gottfried Wilhelm: 15, 113, 204, 208, 299
Lenat, Douglas B.: 117n, 141n
lengua, tener algo en la punta de la, fenómeno de: 221
lenguaje, lengua: 200, 221; de los animales, 52; aprendizaje de la, 80, 82, 112; bioquímico diferente, 197; de computadora, 105-106, 107, 141, 142; desconfianza de los científicos hacia el, 21; español, 71, 82, 107, 118; de la filosofía, 18-21; la heterofenomenología y el, 242; inglés, 20, 74; intencional,

126; las máquinas virtuales y el, 107-108; de las matemáticas, 21; meta-, 18-19; del pensamiento: 51-52, 82-83, 115-117; el significado y, 51-52, 54, 68 , 73, 126; *véase también* gramática; oraciones; palabras
Lenguaje del Pensamiento: 82 lenguas extranjeras: 295 leones: 174
Lepore, E.: 56n
lesión, daño: 279, 282
Lettvin, J. Y.: 185
levantamiento: 161, 166, 167-168, 190
Levine, Joseph: 243, 244 levitación: 251
lex parsimoniae (ley de la parsimonia). *Véase* Ockham, navaja de
leyes: 252; de la física, 65, 66, 120, 121-122, 133, 137, 287; de la física del mundo de Life, 255, 257, 259; de Kepler, 37; el libre albedrío y las, 284; navales, 34n; Segunda Ley de Orgel, 165, 171, 199; de Sturgeon, 36, 49, 289, 299 *libertad de acción, La* (Dennett): 65 libertad: 17, 265, 295
libre albedrío: 14, 22, 44, 50, 51, 57, 114, 251-288, 289, 302; absoluto, 280; contexto causal y, 17, 251-252, 264-265; el determinismo y el, 42, 254-261, 265, 267, 271-278, 279, 284, 285, 286-288; la evolución y el, 26; los hechos históricos inertes y el, 261, 267-271; como ilusión, 251-253; la impredecibilidad y el, 262-265; el incompatibilismo y el, 35, 265; puntos de vista de los científicos sobre el, 251-252, 253, 287-288; la responsabilidad y el, 278-281, 283-284, 285-288; resumen sobre el, 287-288; el sphexismo y el, 281-284; el yo y el, 238-239
libros: 118; en la biblioteca de Babel, 151-156, 165; en el Espacio de Diseño, 161; de números aleatorios, 270-271
Life (modelo de un mundo determinista): 254-261, 269; principios importantes ilustrados por, 261-262
línea de posición: 28n
líneas de descendencia: 161
lingüistas, lingüística: 86, 112, 131, 191, 293 líquido cefalorraquídeo («espíritus animales»): 113n
líquidos, física naïf (o vulgar) de: 292
Lisp sueco: 141, 144 lista del mercado: 118
literatura: 200, 291; *véase también* novelistas, novelas; obras de teatro; poesía, poetas
«Lo siento, pero tu alma acaba de morir» *(«Sorry, But Your Soul Just Died»)* (Wolfe): 85
lobos: 52n, 178, 179, 282 lobotomía prefrontal: 236 lóbulos frontales: 236
loci: 156
Locke, John: 299

lógica: 31, 136, 137, 181, 249, 261, 268, 278, 281, 300
«Logical Geography of Computational Approaches» (Dennett): 80n lógicamente posible: 203-204, 218 lógicos: 108
loterías: 18; imparcialidad de las, 266-267
Lovelace, Ada: 96 lubricantes: 127 lucio: 172
Lucrecio: 266
Lucy (personaje de *Peanuts*): 48
luz: 42, 54, 228; velocidad de la, 102, 155, 189, 261
Lloyd Morgan, Conwy: 37
Mac, computadoras: 91, 105, 106, 159
MacKenzie, Robert Beverley: 160, 166, 170, 171 macrocausal, nivel: 174
Machinery of the Brain, *The* (Wooldridge): 281
Madre Naturaleza: 130, 131, 146n, 165, 173, 186 magia, magos: 221-225; con cartas, 29-30, 31, 222-226, 264; de verdad, 78-79, 84, 222-223, 251 mal aliento: 180
«maldición de la coliflor, La»: 211-213, 304 malignos, genios: 16, 136, 212 malinterpretar: 144, 249
mamífero(s): 29, 51, 159, 182, 280; imposibilidad de los, 175, 280; índice de mutaciones en los, 156; maomeno, 77; Primario, 175-177, 179, 186
manos: 70, 238
mantis religiosas: 52n
manufactura, repetición en: 183-184 manzanas: 77, 265
mapas: 118
máquina de conexión: 146 máquina universal: 99-100, 112 máquina universal de Turing: 106, 260
Máquina Virtual Java (MVJ): 105
máquinas: 204; de movimiento perpetuo, 134; plasticidad de las, 104; reales, 103, 104, 105; el significado y las, 114; versátiles, 103; *véase también* computadoras
máquinas de registro: 85-101, 103, 231, 232; busca, 97-98; como computadoras que almacenan programas, 96, 98-99; copia, 89-91, 96; diagramas de flujo para, 88-91, 93-96, 97; divide, 94, 95-96, 97, 100; ejercicios para, 88, 94-95, 96-97, 305-315; internalizar los procesos básicos de, 90, 91; maomeno y, 96; mueve, 88-89, 96, 98; multiplica, 94, 96, 97; programa de suma no destructiva para, 91-93; resta, 93, 96, 97, 98, 100; RodRego, 91-93; secretos acerca de, 96-99; suma, 87-88, 90-91, 93, 96, 97, 98, 101; tres instrucciones para, 86-87; universales, 100 máquinas expendedoras: 69 ; *véase también* tragacuartos máquinas Linux: 106
máquinas nucleares para hacer palomitas: 165 máquinas virtuales: 76, 100, 102,

103-108, 111, 112, 114, 231; Java, 105-106; jugar al ajedrez y las, 106-107; significado original del término, 104-105; sistemas operativos como, 105-106
mar, determinar la posición en el: 28-29
Marceau, Marcel: 105
margaritas: 159, 160, 182; *Umwelt* de las, 59
Margulis, Lynn: 169
marineros, términos náuticos de: 23
Marioneta, señor: 284-286 marionetistas: 283, 303 mariposas: 263
Marques, J. F.: 214n
Marte, marcianos: 167n, 235; dos cajas negras y, 145-146
Marx, Karl: 44, 75
marxismo: 75
Mary, la científica del color (el argumento del conocimiento): 245-248
masa: 133; centro de, 237, 243 masbienismo: 43-45, 46, 137, 292
matemáticas, matemáticos: 22, 108, 155, 176, 294; herramientas de pensamiento, 15, 20; lenguaje de las, 21; de la mecánica cuántica, 228; pruebas de aleatoriedad en las, 271; soluciones laboriosas en contraste con las fáciles y las, 13
material: 37, 149, 237, 267
materialismo: 32, 33, 203, 236, 246, 287, 302; los genes y el, 80, 205, 207, 301; la mente y el, 77
matrimonio: 179
Matrix, la: 204
máxima del criptógrafo: 132, 133
MCC (Microelectronics and Computer Technology Corporation): 141n
McClelland, Jay: 80n
mecánica cuántica: 37, 228-229, 267; interpretación de los mundos múltiples de la, 42, 137
medio, conciencia como un: 42 medios de comunicación: 107n Mefistófeles: 290, 291 megabyte: 271
meiosis: 196
Melville, Herman: 209, 291 memes: 198-200
memoria: 71n, 72, 107n, 115, 204, 289; del Hombre del pantano, 135, 137; tragacuartos y la, 122; visual, 184n; el yo y la, 238, 239, 240; *véase también* recordar
memoria, cajas de: 106
memoria de computadora: 106, 273; en las dos cajas negras, 138-139, 146; ram (memoria de acceso aleatorio), 101, 269, 270; ROM (memoria de sólo lectura), 98, 104, 112, 159; de «Spakesheare», 189, 190

memoria intermedia: 118
Menabrea, Luigi Federico: 96n
Menón (Platón): 16 mental, competencia: 56 mental, enfermedad: 282 mental, imagen: 206, 233 mentalés: 82
mentales, blockeos y «no cabe duda»: 47 mentales, estados: el robot gigante y los, 128; la Tierra Gemela y los, 123
mentales, representaciones: 51 mentalistas, términos: 126, 142-143
mente: 14, 50, 78, 149, 163, 170, 291; de los animales, 37; como cerebro, 32, 113, 236; el cerebro en contraste con la, 37, 69 , 113n; concebir y la, 205; conciencia y, 201; definición de, 64; descompuesta en submentes, 73; las emociones y la, 76; los homúnculos y la, 73-74, 107; la inteligencia artificial fuerte y la, 226-227; la intencionalidad original y la, 119; de la Madre Naturaleza, 131; metáfora de la computadora para la, 86; el operador «no cabe duda» y la, 47; el operador maomeno y la, 77, 282; las palabras y la, 51-52; la postura intencional y la, 64; la prueba de Turing y la, 229; la psicología vulgar y la, 61, 63-64; los puntos de vista materialista y dualista de la, 32; las razones y la, 172; su reflejo en la evolución, 130; el sistema intencional y la, 64; el sphexismo y la, 282; teoría de la, 60, 61 mentiras: 63n, 125, 128, 238, 253, 288
mercado bursátil: 76 mesas: 57; «dos», 57
metabolismo: 180, 183, 217, 241, 250 metaética: 19
metafísica: 37, 177, 207n, 278, 284; analítica, 293; de la imagen manifiesta, 294
metáforas, lenguaje metafórico: 15, 20, 129, 165; los artefactos y las, 121, 126, 128; de la biblioteca de Babel, 153; del cerebro, 85-86
metalenguaje: 18-19 metales: 134, 189, 202
metas: 60, 63, 67, 129, 149, 299; del autor, 14, 17, 289; de la ciencia comparadas con las del arte, 291; dos cajas negras y las, 144; de historiadores de la filosofía, 299; de humanos comparadas con las de genes, 128-129; el robot gigante y las, 127
meteoritos: 241
método de «primera persona»: 241, 242, 243 método de «tercera persona»: 241
método del escultor, método del autor comparado con el: 65
método socrático de pregunta y respuesta: 233
Meyer, Stephen C.: 38-39
«Mi papá es doctor»: 68 , 78 microbios: 180
microcausal, nivel: 174 microneurocirujano cognitivo: 55-56 microrrobots: 72
microscopía electrónica: 270 microscopios: 243, 273 microtúbulos: 33
miedo: 62, 80
Mies van der Rohe, Ludwig: 190

milagros: 167n, 168, 170, 185, 191, 251, 281, 287
Millikan and Her Critics (Ryder *et al.*): 130n Millikan, Ruth: 121n, 122, 123, 130n, 133, 186 mimos: 105
«Minds, Brains and Programs» (Searle): 226-227
Mind's I, The (Hofstadter y Dennett): 17, 227, 303
Minsky, Marvin: 123 mirar: 211
Misialowski, David: 294
misterio: 149, 150, 185, 201, 214, 250; del ADN, 259; de la conciencia, 148, 223, 251, 301; de Dios, 301; de la mente, 78
«mito de museo» del significado: 133 mitocondria: 180-181
mobiliario: 59
Moby Dick (Melville): 154, 155, 156, 209, 291 modelos «conexionistas»: 80
modelosdeconciencia«orwellianosyestalinistas»:304 modelos neuroinformáticos: 60
modestia: 150
moléculas: 50, 57, 58, 72, 73, 80, 117, 133, 174, 261; estructura de las, 197 molinillo de café: 103 mónadas: 45
Monadología (Leibniz): 299
Monroe, Marilyn: 216 monstruos: 188, 189 montón, los del: 299
moral, moralidad: 22, 44; la evolución y la, 26, 176, 200; el libre albedrío y la, 252, 253, 274, 278, 288; el sphexismo y la, 282
Morgenstern, Oskar: 263
Morse, código: 146 mosca de la fruta: 254 moscas: 159, 185, 186
Moser, David: 233n motivaciones: 238, 286
Motor Analítico: 96
motor de búsqueda con capacidad de operación paralela: 191
motor de combustión interna de gasolina: 183 motor inmóvil: 118
motores semánticos: 133-134, 145
motor(es) sintáctico(s): el cerebro como, 133-134; las computadoras como, 143, 144-145 motores: 183
motosierras: 66, 161
movimiento perpetuo, máquinas de: 134
Mozart, Wolfgang Amadeus: 155, 166, 188, 193 muerte: 139; sin haber tenido hijos, 29, 30, 159, 182; inevitabilidad de la, 44; en la primera infancia, 181; la selección natural y la, 29
mueve, programa para máquinas de registro: 88-89 mulas: 76
muletas de razonamiento: 43-46, 136; apilamiento, 44, 46; la escoba de Ockham, 38-39, 41; la habitación china como, 227-228, 229-234; Mary, la científica del color como, 245-248; masbienismo, 43-44, 46; el operador «no cabe duda» y

las, 46; pasodoble de Gould, 45-46; uso de Gould de las, 43-46; uso del término, 23
mundo: la conciencia visual y el, 248, 249; creación del, 204; descripciones del, 145; determinista, 254-262, 271-278; estar arraigado en el, 133; habitar el mismo, 144, 146; heterofenomenológico, 242; como libreta de un solo uso, 146; real, 194; virtual, 194-196
mundo de juguete: 273-274 mundos reales: 194
mundos virtuales: 194-196, 254
murciélagos: 211, 213; ecolocalización en los, 160, 162
museos: 268, 299
música, músicos: 170, 200, 269-270, 291; artificial, 191, 192-193, 195; canciones, 57, 166, 185, 193, 202, 235, 269, 289; pensar afuera de la caja y la, 41-42; los qualia y la, 211, 213; sinfonías, 161, 166, 193
mutaciones: 26, 42, 156, 161, 177, 182, 205, 261 MVE (Máquina Virtual Española): 107
MVF (Máquina Virtual Francesa): 107
Nagel, Thomas: 38-39, 211n, 213
naïf, física. *Véase* vulgar (o naïf), física
«Naïve Physics Manifesto, The» (Hayes): 292
«Nano-Intentionality» (Fitch): 129n nanodispositivos: 161, 162 nanorrobots: 72, 189, 202
narcisismo: 59
narración en primera persona: 209 naturaleza muerta: 257, 261
naturaleza: 149, 172, 195, 240; ciclos de la, 183; diseño en la, 162-165; estado de, hobbesiano, 17, 154n; interna, 131; leyes de la, 68 , 78, 137; Madre, 130, 131, 146n, 165, 173, 186; el robot gigante y la, 126, 128
Nature via Nurture (Ridley): 158 naval, ley: 34n
navegador: 99
navegantes, navegación: 28
necesidades: 242-243; el robot gigante y las, 125-126, 127 negligencia: 253 neodarwinismo: 163 nervio óptico: 70, 102
Net of Magic (Siegel): 78
Neugebauer, Otto: 187, 188 neurociencia cognitiva: 64, 107n, 284 neurociencia informática: 107n
neurociencia: 33, 243, 253; cognitiva, 64, 107n; informática, 107n; como ingeniería inversa del cerebro, 85
neurocirujana nefanda: 252-253, 282, 283, 285, 288 neurocriptografía: 55-56
neuroestimulación, aparatos de: 283 neurofisiología: 246 neuromoduladores: 32,

107n, 133, 282
neuronas: 33, 55, 75-76, 180, 184, 243; en el área cortical v4, 33; circuitos de las computadoras comparados con las, 102; conciencia y las, 76; efectos cuánticos en los microtúbulos de las, 33; egoístas, 76; como microrrobots, 72; del nervio óptico, 102; como robots biológicos, 75, 202; sinapsis entre las, 32 neuroquímica, economía de la: 76 neurosis: 291
neurotransmisores: 32 neutralidad del sustrato: 109 neutrones: 152
«New Replicators, The» (Dennett): 198
*New York Review of Books:*45-46, 227
New York Times, The: 163
Newton, Isaac: 15, 61, 188, 204, 237, 291
Nichols, S.: 293n nidos: 60, 162, 171
Nietzsche, Friedrich: 299
niño esclavo, enseñar geometría al: 17
niños del Brasil, los (experimento mental): 284-287
niños del Brasil, Los (película): 284
niños, niñez: 55, 143, 156, 161, 202-203, 235; adquisición de la psicología vulgar en la, 61-62; aprendizaje de la lengua en la, 82, 112n; educación de los, 170, 207; estados como de creencia de los, 69 ; la imagen manifiesta y los, 59-60; maomeno y los, 57, 78; muerte de, 181 nivel personal: 238
«no fumar», carteles de: 53
«No te pego si tú no me pegas»: 34n
«no típico de alguien», estado: 238, 239
Nobel, premio: 32, 33, 290
Noche estrellada (Van Gogh): 290 norma, corrección de la: 269
notación científica, números en la: 101 novedades: 41, 68 , 183; diseño y las, 161; las mutaciones y las, 26 novelas: 209-210
novelistas, novelas: 15, 36, 200, 240, 290 nucleótidos: 156
numerales: 172, 173
números: 37, 172-173, 176, 187; aleatorios, *véase* generadores de números aleatorios; de Bernoulli, 96; primos, 173, 301; reales, 101; en registros, lo que representan los, 97, 101
números de Bernoulli: 96n
ObscureCursor: 159
obsesivo compulsivo, trastorno: 252 occipital (visual), corteza: 70, 82, 216
Ockham, escoba de: 38-39, 41
Ockham, navaja de: 37-38, 79 ogros: 264
oído, sentido del: 61, 70, 221, 270 oídos: 270

ojos: 70, 82, 102, 131, 161, 184n, 249, 283; de las ranas, 185, 186, 187
olfato: 57, 61, 70, 211, 221 olvido: 240
ontología: 58-59, 258
operador «no cabe duda»: 46-47, 211, 219, 248, 267 operador maomeno: 77-78, 82n, 117, 147-148, 174, 176, 268; el ajedrez y el, 273, 274; el cerebro y el, 114; la conciencia y el, 229-330; las creencias y el, 57, 77, 142, 148, 208; las dos cajas negras y el, 141, 142, 148; los elevadores y el, 112; las especies y el, 178-179; la evolución virtual y el, 194-195; la física cuántica y el, 228-229; las máquinas de registro y el, 96; la mente y el, 77-78, 282; los programas OCR (de Reconocimiento
Óptico de Caracteres) y el, 98; las proteínas y el, 261; el tragacuartos y el, 119-122; opiniones del robot gigante: 128
oportunidades: 57; errores como, 26, 27-28, 30; el libre albedrío y las, 271, 273-274, 277; reales, 17
oraciones, enunciados: 60; en la biblioteca de Babel, 153; dos cajas negras y, 140, 141, 142, 144-145, 146; genes comparados con, 158; nuevas, entenderlas, 68 ; significado y, 53, 54, 64n, 115-117, 118, 140, 141-142, 144-145, 146, 147; verdaderas y falsas, 146
organismo(s): 157, 260; diseño de, 161, 162, 165, 167, 169, 170, 171; la especiación potencial y el, 179, 182; «modelo», 254; ontología (*Umwelt*), 59; razones de los, 182; unicelular (mónada), 44
Orgel, Leslie: 165
Orgel, Segunda Ley de: 165, 171, 199
original (o intrínseca), intencionalidad: 118-119, 121, 123-125, 128, 129, 131, 146-147, 148 originalidad: 291
oro: 267-268 osos: 31, 162
osos hormigueros: 59 osos polares: 162
Otto, ciclo de: 183
Otto, Nikolaus: 183 ovarios: 156
Oxford English Dictionary: 166
Oxford: John Locke Lectures en, 21; tesis de doctorado del autor en, 20
padres: 27, 156, 159, 161, 180, 182; adoptivos, aves como, 171
palabras: 82, 176, 199; amor como, 48-49; desconfianza de los científicos hacia las, 21; los genes como, 157-159; como herramientas de pensamiento, 19-20; la imagen manifiesta y las, 59; la intencionalidad original y las 123-125; leídas por la computadora, 100; lista de, en orden alfabético, 108; procesadores de, 99, 141n, 152n, 169, 189, 270; el significado y las, 51, 54, 59-60, 118, 123-125; en el vocabulario, 158 palomas: 22n; el significado y las, 54

pan comido: 19 panales: 173
Panamá, el tragacuartos y: 120-123, 128, 130, 137, 147
papeles, el yo y los: 239-240 parásitos de puesta: 171
Parménides: 38
parodia: 34; de razonamiento, *véase* reducción al absurdo
párpados, forma de los: 29
parsimonia, ley de la *(lex parsimoniae). Véase* Ockham, navaja de partículas: 152, 153, 254, 276 pasado virtual: 190
pasado: 267, 277; mundo de Life y el, 261; preguntas sobre el, 240 pasarse de listo: 33
Pascal, Blaise: 15, 113
«pastor mentiroso, El» (fábula de Esopo): 17 patines submarinos: 165
pato, ver a un: 82
Pauli, Wolfgang: 30, 292
Pauling, Linus: 291
PC (computadoras personales): 91, 105, 106, 107 PDP Research Group: 80n
Peanuts, tira cómica: 48 pecado original: 27 pececito: 65
peces: 37, 65, 83, 124, 159, 160, 182, 254, 274, 282 peculiaridades: 71
pegajosidad: 19
peligrosa idea de Darwin, La (Dennett): 20n, 34n, 108
«pelirrojos, algo con los»: 115-117
pelotas, rodando hacia abajo en una colina: 62 péndulos: 62
Penfield, Wilder: 203
Penn y Teller: 78
Penrose, Roger: 33
pensamiento, herramientas del: 14-23, 24-50; adquirir o hacer, 19-20; antropomorfismo como, 165; como armas, 20, 49; computadoras como, 84-114; para la conciencia, 201-250; diseño de, 161-162; elaboración de, 14-15; para la evolución, 149-200; para fines especiales, 14; informales, 20; lista de Hofstadter de las, 19-20, 60; malas, 14, 22-23, *véase también* muleta de razonamiento; matemáticas, 15, 20; palabras como, 19; el punto fijo y las, 22; resumen sobre las, 49-50; para el significado o contenido, 51-83, 115-148; simples y generales, 14 pensar afuera de la caja: 20, 41-43, 61
pensar, pensamiento: 116, 199, 201, 248, 265; sobre el ajedrez, 273, 275; de dos cajas negras, 147; desiderativo, 50, 201, 236, 288; es difícil, 13-14; intencionalidad y, 118; lenguaje del, 52, 82-83, 115-117; maomeno, 273; ponerse meta-y, 18; previo, recuerdos del, 27; reflexión sobre el, 27, 30; del robot gigante, 127; el significado y el, 130-131, 147 pentágono: 206

pequeña locomotora que sí pudo, La: 202-203
per (Programación Ensambladora de Registros): 87, 91-95, 96-97
percepción: 76, 80, 130, 184, 204, 282, 289; la caverna de Platón y la, 17; la conciencia y la, 221, 236, 243; las reacciones negativas y la, 226; el significado y la, 51, 115; el tragacuartos y la, 121, 122-123; el yo y la, 236
perros: 37, 38, 58, 159, 175, 178-179, 182; entendimiento de los, 51; maomeno, 77; l significado y los, 51, 54
perros chihuahueños: 182 persecuciones inútiles: 19, 176 personal/subpersonal, distinción entre: 69 -73 personas: 69 -73; como cerebro, 70-71; creencias reales de las, 69 ; enfoque de arriba abajo sobre las, 73; partes subpersonales de las, 69 -73; la postura intencional y las, 64, 69 ; como sujeto de historial, 72
perspicacia: 191, 192, 283
Persuasión (Austen): 209-210 peso: 180
Peterson, Dale: 184n
petróleo, programa de computadora para refinería de: 67
pez cebra: 254
Philosophiae Naturalis Principia Mathematica (Newton): 291
Philosophical Foundations of Neuroscience (Bennett y Hacker): 76n
Philosophical Transactions of the Royal Society: 284 philpapers.org: 47
Piatelli-Palmarini, M.: 130n pie de atleta: 76, 180
piedra, papel y tijeras, juego de: 262-263, 265, 267, 278
piedras, rocas: 65, 184 piel, células de la: 157 piernas: 34, 161
pies de barro: 19
Pinker, Steven: 83n pintura, colores de: 58 pinturas: 118, 161, 218, 299 piratas del espacio: 146 pixeles: 260, 261
placas tectónicas: 250 planeación: 128, 238
planeadores: 258, 259, 260, 261, 262
planetas: 137, 149, 172, 197; órbitas de los, 37; la Tierra Gemela, 123-125, 137
Planiverse (Dewdney): 260n
plantas: 57, 68 , 126, 254; evolución de las, 161; genoma de las, 157
Plantinga, Alvin: 304n
plasticidad: del cerebro, 71; de las computadoras, 106; de las máquinas, 104
Platón: 73, 299
Platón, la caverna de: 16, 17 playas: 165
poblaciones: 182n; aislamiento de las, 177-178; variaciones en las, 171 podadoras: 103, 161
«podría haberlo hecho de otro modo»: 278 poesía, poetas: 20, 22, 58, 128, 161, 165, 235, 290; sonetos, 41, 161 polígonos: 206

«Politburó», control de: 75
«ponerse meta-»: 18, 19, 200
Popek, Gerald: 104
Popper, sir Karl: 32; su zambullida en el Gran Canal, 32n
Population Thinking and Natural Selection (Godfrey-Smith): 304
póquer: 262
póquer, cara de: 262, 264 pornografía: 22n
portátiles, computadoras: 13, 85, 86, 99, 100, 104, 106, 231, 260, 270; copiar en, 269
Porter, Cole: 202
posibilidades en el ajedrez: 273-274
posible: en el ajedrez, 273; lógicamente, 203-204, 218 postimágenes: 213
postura de diseño: 65-67, 147, 148, 161, 163; el mundo de Life y la, 258, 259, 261
postura física: 65, 67, 147, 148; el mundo de Life y la, 257, 258, 259, 261
postura intencional: 63-70, 77, 116, 174, 263, 274; la conducta y la, 172; definición de, 64; las dos cajas negras y la, 146, 148; la heterofenomenología y la, 242-243
Poundstone, William: 258n, 260n
PowerPoint: 93
precio de salida, en anuncios: 264
predicción: 263; en el ajedrez, 273; el centro de gravedad y la, 237; las dos cajas negras y la, 140, 142-143; en Life, 257, 259, 261; desde la postura de diseño, 65-66, 67; desde la postura física, 65-66, 67; desde la postura intencional, 63-64, 66-68; el tragacuartos y la, 120 preferencias: 71, 83; de color, 217-218 pregunta(s): acerca de la conciencia, 221; en el debate sobre la evolución, 166; favorita de Feynman, 254; grandes, y los científicos, 22; metafísicas y epistemológicas, 177; sobre el pasado, 240; plantear, 25-26, 51; retóricas, 47, 52n, 137
presentación de la persona en la vida cotidiana, La (Goffman): 239 presión sanguínea: 217
prevenir, previsiones: 83, 85, 265; el cerebro y las, 83, 85, 133; el robot gigante y las, 125, 127 previsión: 129, 183, 190-191
Primario, Mamífero: 175-177, 179, 186 primates: 29, 51, 182
primos, números: 173, 301
Princess Bride, The (película): 263
Principia (Newton): 188
Principia Philosophiae (Descartes): 291 principios: 115
probabilidad: 257, 285, 286
problemas: comienzo de los, 42; la conciencia y, 220-221; interesantes, 141;

soluciones creativas a los, 141; picar, 65
problemas de juguete: 75n, 110, 254 problemáticas filosóficas: 296 procariotas, células: 129n
procesamiento, unidad de: central (cpu), 139, 231, 273; máquina de registro, 86, 87, 91, 98, 99 procesos informáticos: 76, 80
procesos oponentes: 76 profesores: 128 profulidad: 48-49
programa de copiado para máquinas de registro: 89-91
Programador Maestro: 114
programadores de computadora: 108, 110-112, 116, 184, 276; la primera programadora, 96n programas de computadora: 79, 84, 85, 86-103, 104-108, 193n, 196; como algoritmos, 109; código fuente de, *véase* código fuente; compilador, 108, 111-112; de las dos cajas negras, 138-139; fotografías y, 98; la inteligencia artificial fuerte y los, 226, 299-232; la máquina virtual como, 104-106; para máquinas de registro, 86-99; para números aleatorios, 270-271; la prueba de Turing y los, 229-230, 233; para refinerías de petróleo, 67; seudoentendimiento de los, 281; subrutinas, 158
programas integrados: 100, 101, 103, 104
progreso: 148; el apilamiento y el, 44; errores como clave para el, 25; las grandes preguntas y el, 22; pensar afuera de la caja y el, 42
promesas: 54, 58, 60, 63n prominencia: 259 pronósticos del tiempo: 107 propiedad: 268 proporcionamientos: 59
proposiciones: 31, 56, 116, 148, 160, 292, 294; dos cajas negras y, 144; profulidad, 48-49
propósito: 130, 137, 147, 149, 162, 163, 171, 186 prosa: 20
prosopagnosia: 215-216, 218, 220 proteínas motoras: 72, 202
proteínas: 129n, 134, 160, 174, 184, 189, 195, 196, 228; diseño de, 261; como nanorrobots, 72
protones: 152, 269 proyecto Cog: 117n proyectos: 137, 238
prueba de concepto: 75, 108 psicofísica: 243
psicología: 116, 253, 292n, 300; académica o científica, 60; animal, 69 ; cognitiva, 243; contrastada con la neurociencia, 107n; cursos de introducción a la, 62; experimental, 80; proyecto Cog y la, 117n; vulgar, 52, 60-65, 68 , 130, 210
psicología cognitiva: 243 psicología experimental: 80 psicología Gestalt: 60 psicólogos: 86, 214, 300 psiquiatras: 239
público lego como señuelo: 39-41
Publisher's Clearing House: 266
Puccini: 193 pudúes: 55, 115 puentes: 57, 161
pulmones: 85, 130, 161, 246 pulsaciones por minuto: 217 puntos de ramificación:

266 puntos fijos: 22, 65
Putnam, Hilary: 123n
qualia: 211-220, 304; cambio en sentimientos y los, 219; color, 211, 217-220, 247-248; la comida y los, 211, 218, 219; cuatro condiciones de los, 212; definición de, 211-212, 213; error que se comete con los, 211
quark y el jaguar, El (Gell-Man): 228-229 quarks: 57
quiliógono: 206, 247
química, químicos: 25, 58, 117, 260, 287, 300
Quine, W. V. O.: 19n, 131-133, 145n, 152n, 153n
«Quining Qualia» (Dennett): 304
racional: 73, 176-177, 285 racionales, agentes, 64, 72, 77
racionalidad: 37, 56, 61, 150; la postura intencional y la, 64, 66-69, 242, 243
racionalismo: 299 radar: 162
radiación electromagnética: 58 radiactvidad: 54
ram (memoria de acceso aleatorio): 101, 269, 270 ramificación condicional: 96, 99, 114
ramificar: 87; *véase también* instrucción disminuye o ramifica
ranas: 159, 185-187 rand Corporation: 271
Randi, James: 78, 79, 80
Ranger (el robot que más ha caminado): 164
Rapoport, Anatol: 34, 35n Rapoport, reglas de: 34-35, 36, 49 rastreo del cerebro: 134
rastrillos espaciales: 258 ratas: 242, 254
ratón: 158 ratonera: 113, 161
Rawls, John: 300
razón de ser: 129, 130, 169, 293 razón: 50, 163
razonable: 60
razonamiento: 18, 64, 67, 143, 160, 170, 179, 236, 263; deductivo automatizado, 292n; «extraña inversión» del, 231; por parodia de, *véase* reducción al absurdo
razones de libre flotación: 171-172, 173, 174 razones: 149, 201; para el diseño, 163-164; la naturaleza y las, 146n; la selección natural y las, 130, 163-164; *véase también* razón de ser reacciones afectivas al color: 247
«Real Patterns» (Dennett): 143n
Realidad Virtual: 17, 204
realidad: 149, 191, 211; en el ajedrez, 275; la caverna de Platón y la, 17; de la conciencia, 251
realismo histérico: 150, 151, 177
recetas: ADN como, 26, 157; algoritmo como, 109; genes como, 196

reconocimiento facial: 215-216, 217, 220
Reconocimiento Óptico de Caracteres, software para (OCR): 98
reconocimiento: 191, 220, 240, 263; de color, 219, 220; facial, 215-216, 217, 220; el Hombre del pantano y el, 135; en el robot gigante, 126; en la selección natural, 130; en el tragacuartos, 130
recordar: 76, 77, 85, 241; pensamiento previo, 27; síndrome de Korsakoff y, 56; *véase también* memoria
red de conmutadores telefónicos, sistema nervioso como: 73
red de mate: 273
Red Queen, The (Ridley): 170
«redes neuronales», modelos de: 80
reducción al absurdo: 16, 31-34, 123, 175, 230; pasarse de listo y la, 33; las preguntas retóricas y la, 47-48; la sobresimplificación y la, 33
referencia: 134, 146 reflejo: 113n
reflexión: 27, 30, 160, 200, 201, 236, 249 refrigeradores: 104
refuerzo, aprendizaje por: 27
registros: 86-96, 97-99, 103; acumulador, 101; cantidad de, 101; contenidos de, 86-93, 94, 95, 97; dirección de, 86, 88, 97, 98, 101; en el *hardware,*106, 107; memoria intermedia, 91; reducir a cero, 89, 90, 91, 93, 94, 95
regresión: finita, 74, 129n; infinita, 73, 74, 176 regresión finita: 74, 129
regreso infinito: 73, 74, 176
regularidades: 262; en el ajedrez, 272, 273, 274; motivos en pixeles, 260 regularidades, reconocimiento de: 80 relaciones formales: 179
religión: 150, 199, 304; el diseño inteligente y la, 162; el masbienismo y la, 44
relojes: 183, 243; despertadores, 65, 66; dos cajas negras y, 140
repetición: cíclica, 183-185; en la manufactura, 183
Réplica de Sistemas: 230
representaciones: 118, 129, 130; del ajedrez, 273; de un mundo común, 145
reproducción: 167, 198-199, 200, 259 reproducción: 183, 185, 196-198, 250, 254; el
aislamiento y la, 177-178; asexual, 169; de los chapulines, 173; sexual, 169-170, 173, 200
reptiles: 159, 175, 182, 280 *república, La* (Platón): 16 resfriado común: 199
responsabilidad: 238, 252, 253; absoluta, 278-281; el libre albedrío y la, 278-281, 283-284, 285-288; máxima, 278-281; moral, 276, 278-281, 288 resultados garantizados, de algoritmos: 109 resultados negativos: 30n
retina: 228, 246; neuronas sensibles al rojo de la, 33; de la rana, 185
retóricas, preguntas: 47-48, 52n, 137 retroalimentación: 19, 60, 283

retroalimentación, bucles de: 263
rey Lear, El (Shakespeare): 290
Ridley, Matt: 156n, 158, 170
riesgo: 192, 259, 277, 286; de la postura de diseño, 66; de la postura intencional, 66; de la reproducción sexual, 169
risc (computadora de conjunto de instrucciones reducidas): 101
Robert de Chester: 108
«robot, atrapado en el centro de operaciones del»: 81-83
robot gigante: 82, 125-129, 147, 148
robótico, sistema de conducción de carros: 61 robots: 60, 61, 162; que caminan, 164-165; la conciencia y los, 202, 204, 221-222; diferencia entre la calculadora y los, 77; dos cajas negras y los, 147; la física vulgar y los, 292; flotante, 259; gigantes, 82, 125-129, 147, 148; proyecto Cog, 117n; los qualia y los, 213, 218; responsabilidad por los, 279; sistemas de control, 119, 127, 128; «Spakesheare», 189 robots húmedos: 44
RodRego: 91-93
rojo: conciencia del, 33; los qualia y la, 212
rom (memoria de sólo lectura): 98, 104, 112, 159
Romeo y Julieta (Shakespeare): 290
Romper el hechizo (Dennett): 304
Ross, Amber: 207n
Rubens, Peter Paul: 193
ruido: 194-196, 257, 261, 269
Ruina, Andy: 164 ruleta: 271
Rumelhart, David: 80n
Russell, Bertrand: 36, 299
Ryder, Dan: 130n «información clasificada», principio de: 171 sacar el promedio: 19
Sagrada Familia, La: 172
Salinger, J. D.: 209
saltación, la evolución y la: 43, 187 salvación: 265
sándwich de jamón: 104
Sanford, David: 175
sangre: 113, 130; células de la, 157
Santa Claus: 64n Sartre, Jean-Paul: 36 sastres: 14-15 satélites: 167n, 197
Science and Religion (Dennett y Plantinga): 304
Schank, Roger: 32
Schönborn, Christoph: 163

Schooler, Jonathan W.: 253n
Schubert, Franz: 193
Searle, John: 118, 119, 120, 121, 123, 130; la habitación china de, 16, 205, 226-234
selección cultural: 199
selección natural: 26, 29, 148, 199; cerebro y la, 83, 112; competencia de la, 83; competencia sin comprensión en la, 83, 170-174; el diseño y la, 29, 83, 129-130, 160-170, 172; como algo que encuentra razones, 130, 163-164, 171-172; la evolución virtual y la, 194; la exaptación y la, 186; la imagen manifiesta y la, 59; como la mejor idea, 149; poder de la, 149; subestimar la, 162
Sellars, Wilfrid: 57, 58, 215, 219, 220
semántica: 19, 70, 134n, 190n, 198; dos cajas negras y, 142-143, 146, 147, 173-174
semanticidad: 133, 146 semillas o esporas: 126
sentido común: 229, 279, 284, 287
Sentido de la Vida: 149
sentidos: 72, 113, 133, 146; *véanse también sentidos específicos*
señales, señalizar: 51, 76, 184, 195; el cerebro y las, 113-114; costosas, 174 serendipia: 195, 196
series aleatorias: 262-263, 267, 271 seudoaleatoriamente: 261 seudoaleatoriedad: 140, 141, 272
seudoaleatorios, números, generador de: 271, 273, 275, 276 seudociencia: 162 seudocódigo: 111 seudoentendimiento: 281 seudomemorias: 190
Seung, Sebastian: 76
sexual, reproducción: 169-170, 173, 200 sexual, selección: 46
sexualidad, relaciones sexuales: 22n, 291; como grúa, 169-170, 200
Shakespeare, William: 161, 166, 196, 290
Shannon, Claude: 113
Shawn, Wallace: 263n
Sherwin-Williams: 58
Shklovski, Viktor: 61n
Siegel, Lee: 78, 222
Signature in the Cell (Meyer): 38-39
significado o contenido: 14, 22, 50-83, 196, 289, 302; las características compartidas y el, 53-54; el cerebro y el, 51-56, 62, 69 -72, 109-110, 113, 114, 133-134, 147; comenzar con el, 51; detección del, 54-55; de los genes, 129; los hechos más profundos y el, 130-131, 133, 147, 150; herramientas para pensar acerca del, 51-83, 114-148; la intencionalidad y el, 53-55, 64n, 118-125;

maomeno, 134, 148; materia unida con el, 149; «mito de museo», 133; resumen sobre el, 147-148; ser acerca de algo y el, 53, 64n, 118-119; el sesgo antropocéntrico y el, 54; teoría del, 148; teoría evolutiva relacionada con el, 26, 130, 147-148, 186, 187
significadores insignificados: 118, 129 silicio: 75, 134, 202; chips de, 103, 104, 189
sillas: 57, 59
simbiosis, simbiontes: 169, 180, 200
símbolos: 118, 214, 269; habitación china y, 230-232, 233 simios: 51, 52n, 77
Simmel, Mary-Ann: 62 simpleza de algoritmos: 109 simplicidad: 109, 254
Sims, Karl: 195 sinapsis: 32, 113
síndrome de Asperger: 63n síndrome del sabio: 171 sinergia: 168, 169
Singer, Wolf: 251
sintaxis: 134n, 143, 144, 190n síntesis proteica: 129
sirenas, avistamientos de: 243
«sistema B» babilónico de efemérides lunares: 187 sistema consciente, en el cerebro: 216
sistema experto: 141
sistema inconsciente en el cerebro: 216
sistema intencional: 67-69, 72, 129, 144, 263, 264, 278; definición de, 64; el tragacuartos y el, 119-120
sistema límbico: 216
sistema nervioso: 76, 246; ciclos en el, 184; el masbienismo y el, 44; microtúbulos en el, 33; los procesos oponentes y el, 76; como red de conmutadores telefónicos, 73
sistema sensorial del robot gigante: 126
sistemas de control: 67, 207; del robot gigante, 127, 128
sistemas operativos: 105, 106 sistemas solares: 160
Skinner, B. F.: 27, 291
Smart, J. J. C.: 33
Smith, George: 86n
Smith, John Maynard: 170
Snow, C. P.: 290 sobornos: 63n sobresimplificación: 33
Sócrates: 16n, 176
*software:*38, 231; del cerebro, 71n, 112; de computadora, 75, 84, 99, 100, 105-106, 107n; subrutinas, genes como, 157-159
Sohu Weibo: 232
Sol, luz solar: 68 , 125, 163 solidez: 57, 252
solteros: 60; casados, 301

sombras chinescas javanesas, dramatización para: 303
sombrero de tres picos: 28n sonar: 162
sonido: 42, 54, 211, 221, 244 sonrisas: 63n
Sony: 58
sorites, argumento (o «paradoja del montón»): 280
Spamlet (Spakesheare): 188-191 sphexismo: 20, 281-284
Stanford EnCYClopedia of Philosophy: 203n, 280
Stoneking, Mark: 180
Strawson, Galen: 278-280
Sturgeon, Ley de: 36, 49, 289, 299
Sturgeon, Ted: 36 subastas: 264 subibajas: 147
subjetividad: 59, 241, 242, 243; los qualia y la, 211-212, 219, 226
subrutinas de caja de herramientas: 158-159 sueño: 17, 18, 184
sueños: 51, 241
suerte: 42, 171, 182, 267, 275
sujetos que se representan las razones: 171
summum bonum: 129
supercomputadoras, en las dos cajas negras: 138-139, 140
Supermán: 31, 167, 304
supervivencia, máquinas de: especies como, 128; el robot gigante como, 129-130
supervivencia: 164, 181, 198, 205, 147 superyó: 73
suposiciones, falsas: 42
«Swamp Mary and RoboMary»: 304
Syntactic Structures (Chomsky): 291
tacto: 70
talento: 241, 297; de células eucariotas, 129n; para el color, 218; física vulgar como, 60; genéticamente heredado, 59; psicología vulgar como, 60; del tragacuartos, 130
teatro, obras de: 161, 167, 196, 200, 290; artificiales, 188-191
técnicas de recolección de datos: 295 tecnología informática: 64
tejido maravilloso: 78-81, 83, 134, 222, 251; las computadoras y el, 102, 108
tejones: 51 telarañas: 161 telas tejidas: 162
teléfonos inteligentes: 13 telescopios: 167n teletransportación: 234-235
televisión: 202, 246; a color, 58; control remoto de la, 104
temperatura: 68 -69, 133
«Tengo hambre/J'ai faim»: 53 teoría de cuerdas: 207
teoría de juegos: 263, 283 teoría de la bicicleta: 60
Teoría de la justicia (Rawls): 300 teoría de la mente: 60, 61

teoría de la nutrición: 60
teoría de la probabilidad: 15, 148n, 263 teoría de las señales costosas: 174
teoría política: 300
teóricos de la conspiración: 39 terápsidos: 175, 280
tercera persona omnisciente, narración en: 209-210
Tercera Sinfonía (Brahms): 165 términos masa: 59
términos náuticos, falsas definiciones de: 23 términos: 146; definición de, 176
termitas: 52n; termitero catedral, 172 termostatos: 60, 69 , 144
tesis: 21 testículos: 156
Thatcher, Margaret: 216
Tiananmen, masacre en la plaza de: 233 tiburones: 136, 137
tiburones vaca: 136-137 tics: 59
tiempo: 37, 160, 163, 179, 181, 189, 191, 192, 238, 252, 294; en el mundo de Life, 255, 257, 260; en las partidas de ajedrez, 272, 273
tiendas de abarrotes, dependientes de: 69
Tierra: 163, 234-235, 288; centro de la, 237, 238; rotación de la, 167n
Tierra Biomórfica: 156
Tierra Gemela: 123-125, 135, 137 tigres: 128, 282
*Time's Literary Supplement:*38-39 timina (T): 129, 156, 157, 196-197 tipo de cambio, paridad: 121, 213-214
tipográficos, errores (erratas): 26, 153, 156, 189-190
tipos impredecibles: 19, 60
toma de decisiones: 72, 128, 252-253, 264, 265, 287; del ajedrez, 272-273, 274; maomeno, 274 toma y daca, estrategia del: 34
Ton Beau de Marot, Le (Hofstadter): 194 topes para puertas: 122
torneo del dilema del prisionero: 34
torneos: 262, 267; de ajedrez, 272-275; del dilema del prisionero, 34
trabajadores de cuello blanco: 72
Trabajadores: 73 tradición oral: 166
tradición: 149, 150, 165, 200, 236, 265; artística, 41; conciencia y, 201; oral, 166
traducción: 38, 117, 152, 187-188, 282, 303; de dinero de verdad, 214; de lenguajes bioquímicos, 197; radical, 131-133, 145n traducción radical: 131-133, 145n
tragacuartos: 119-123, 137, 147, 186; la evolución comparada con el, 130; el robot gigante comparado con el, 126, 128
tragacuartos ambulante. *Véase* tragacuartos tragones: 258, 260, 261
transductores: 112, 119, 120 transferencia tecnológica: 168
trastorno de identidad disociativo (antes conocido como trastorno de personalidad

múltiple): 239-240
trato: 58, 60; faustiano, 290-292 trenes de vapor: 258
triángulo plano: 301
Trivers, Robert: 45
Tucson II: 33
Tufts: 86; estudiantes de licenciatura en, 40; Estudio de *Software* Curricular en, 86n, 93
turco: 82
Turing, Alan: 13, 77, 99-100, 106, 113-114, 148, 205, 229-231, 259; y los algoritmos, 108-109
Turing, prueba de: 229-230, 233
Turing, máquina de: 13, 85, 99-100, 103, 106, 260
Turing, máquina de, cuello de botella: 100
Two Cultures (Snow): 290
Tyson, Mike: 216
umbral de fusión de parpadeo: 59, 263
Umwelt: 58-59 una verdad: 132 unicornio: 49
«Universal Library» (Quine): 153
Universidad de Harvard: 298; teatro Loeb en la, 303; Sociedad de Miembros de la, 235
universos: 150n, 196, 204, 212; del ajedrez, 275; la biblioteca de Babel comparada con los, 152, 155; deterministas, 269; dios como creador de, 37; infinidad de, 137; partículas en los, 152, 153
utf-8 (Formato Universal de Transformación de conjuntos de caracteres -8 bits): 141n
utilidad esperada: 263 uvas verdes: 19-20
vacas: 136, 175
Vale of Laughter, The (De Vries): 154n valentía: 62, 176
valor económico, intrínseco y derivado: 214
valor: 150, 267; del dinero, 213-215; relativo de las piezas del ajedrez, 273
Van Gogh, Vincent: 290 *vaporware:*107n variables: 135, 190n
velocidad de caída de cosas pesadas y de cosas ligeras: 16
Venecia: 32; zambullida de Popper en el Gran Canal de, 32n
Venter, Craig: 151, 196 ver. *Véase* vista; visión
verdad: 13, 14, 43, 50, 134, 137, 148, 242, 243, 290, 291, 292, 294, 299; *a priori,* 295-296; las dos cajas negras y la, 140-145, 147; los hechos históricos inertes y la, 268; de orden superior del ajedrez, 294; de orden superior del ajetrez, 294-298; la profulidad y, 48 verdad, hacedor de: 176 vergüenza: 27

víboras: 160, 165
Victim of the Brain (película): 303
vida: 108, 139, 162, 176, 185; arrasamiento de la, 288; como algo asombroso, 160; origen de la, 38-39, 129, 159, 185, 261, 301, 302; santidad de la, 150; Sentido de la, 149; tenacidad de la, 160
Vidal, Gore: 26
video de la animación de Heider y Simmel: 62 videograbaciones: 54, 268, 271
videojuegos: 194 vim: 214, 215 virtud: 41, 176, 301
virus: 172, 177, 180, 181
visión, sistema visual: de color, 213, 217-220, 245-249; de la rana, 185-186, 187; del robot gigante, 126; *véase también* vista visión ciega: 70
vista: 58, 61, 70, 102, 128, 131, 221, 228
visual, corteza. *Véase* occipital (visual), corteza vitalismo: 213, 214-215
Vohs, Kathleen D.: 253n
Volición. *Véase* elección; libre albedrío
Volkpsychologie: 60
Von Neumann, cuello de botella de: 100, 102
von Neumann, John: 13, 14, 20n, 101, 113, 259-260, 263
Von Neumann, máquina de: 13, 85, 260 von Uexküll, Jakob: 58n, 59
vulgar (o naïf), física: 60-61, 292, 294; sofisticada, 292
vulgar, ideología: 251
vulgar, psicología: 52, 60-65, 68 , 130, 210; el autismo y la, 63n; funcionamiento de la, 63-64; al nivel de la persona, 69
vulgar, teoría: 248
Watson (inteligencia artificial): 54-55, 141n, 144
Watson, James: 33, 205, 259, 291
Wegner, Daniel: 251
Weizenbaum, Joseph: 48
Weizenbaum, Miriam: 48
Wetware (Bray): 184
What Darwin Got Wrong (Fodor y Piatelli-Palmarini): 130n
What Do You Care What Other People Think? (Feynman): 14
«What Is It Like to Be a Bat?» (Nagel): 213n
«What RoboMary Knows» (Dennett): 304
«What the Frog's Eye Tells the Frog's Brain» (Lettvin): 185
«Why You Can't Make a Computer That Feels Pain» (Dennett): 304
Wiener, Norbert: 113
Wiggins, David: 277

Wikipedia: 117n
«Will to Believe, The» (James): 24
Wilson, Allan: 180
Williams, George: 196
Williams, Rowan, arzobispo de Canterbury: 49
Wimsatt, William C.: 59n
Windows: 105, 107, 159
Wittgenstein, Ludwig: 36, 61n
Wolfe, Tom: 253
Wooldridge, Dean: 281
Wright, Robert: 223, 251
Wundt, Wilhelm: 60
yeso, para brazo fracturado: 105 yeso virtual: 105
yo, el: 235-241; biografía, «telón de fondo» y proyectos de, 238, 240; como centro de gravedad narrativa, 236-241; interpretación de papeles y, 239-240; responsabilidad de, 279; trastorno de identidad disociativo y, 239-240; ubicación del, 236, 237
Young, Andrew: 217
Young, J. Z.: 20
«You're the Top» (canción): 202
Zahavi, A.: 174n
zimbos: 207-210, 226, 229, 236
zombis: 136, 236; filosóficos, 203-210, 222
«zorra y las uvas, La» (fábula de Esopo): 19-20

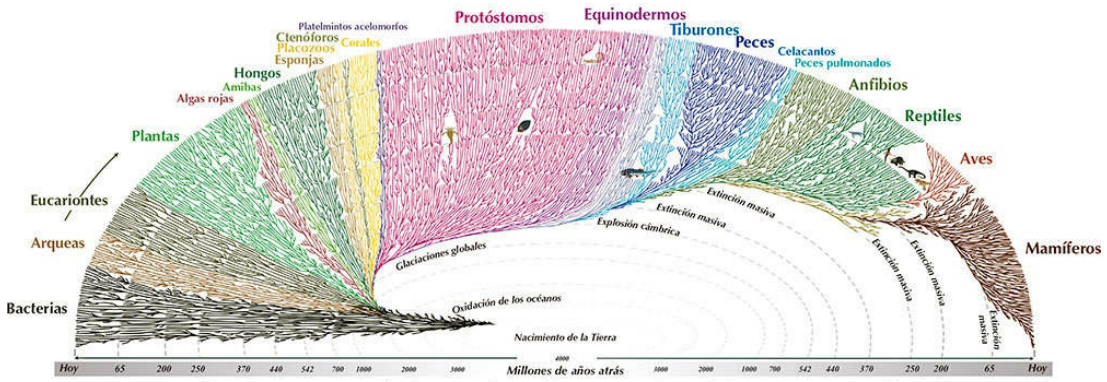


FIGURA 13. *En este diagrama se muestran todas las ramas de la vida mayores y algunas de las menores presentes hoy en la Tierra, y sólo algunas de las que se han extinguido. Por ejemplo: dinosaurios: extintos.*
© Leonard Eisenberg. Todos los derechos reservados. evogeneao.com.

FIGURA 14. *Termitero catedral (Fiona Stewart) y* La Sagrada Familia *(cortesía de* Diario di Viaggio*) (véase la p. 172).*

Índice

Sumario
Prefacio

I. *Introducción. ¿Qué es una bomba de intuición?*

II. *Doce herramientas de pensamiento generales*
1. Cometer errores
2. «Por parodia de razonamiento»: el uso de la *reductio ad absurdum*
3. Las reglas de Rapoport
4. La ley de Sturgeon
5. La navaja de Ockham
6. La escoba de Ockham
7. Usar público lego como señuelo
8. Pensar afuera de la caja
9. Tres especies de *gouldeo*: masbienismo, apilamiento y el pasodoble de Gould
10. El operador «no cabe duda»: un *blockeo* mental
11. Preguntas retóricas
12. ¿Qué es una profulidad?
Resumen

III. *Herramientas para pensar acerca del significado o contenido*
13. Asesinato en la Plaza de Trafalgar
14. Un hermano mayor que vive en Cleveland
15. «Mi papá es doctor»
16. Imagen manifiesta e imagen científica
17. Psicología vulgar
18. La postura intencional
19. La distinción personal/subpersonal
20. Una cascada de homúnculos
21. El operador *maomeno*
22. Tejido maravilloso
23. Atrapado en el centro de operaciones del robot

IV. *Intermedio sobre computadoras*
24. Los siete secretos del poder de la computación revelados

25. Máquinas virtuales
26. Algoritmos
27. Automatizar el elevador
Resumen

V. *Más herramientas sobre el significado*
28. Algo con los pelirrojos
29. El tragacuartos ambulante, la Tierra Gemela y el robot gigante
30. Traducción radical y un crucigrama quineano
31. Motores semánticos y motores sintácticos
32. El Hombre del Pantano conoce a un tiburón vaca
33. Dos cajas negras
Resumen

VI. *Herramientas para pensar acerca de la evolución*
34. El ácido universal
35. La biblioteca de Mendel: Extremada y Evanescente
36. Los genes como palabras o como subrutinas
37. El árbol de la vida
38. Grúas y ganchos celestes: levantamiento en el Espacio de Diseño
39. Competencia sin comprensión
40. Razones de libre flotación
41. ¿Entienden los chapulines los números primos?
42. Cómo explicar los saltos de las gacelas
43. Cuidado con el Mamífero Primario
44. ¿Cuándo ocurre la especiación?
45. Hacedores de viudas, Eva Mitocondrial y coronaciones retroactivas
46. Ciclos
47. ¿Qué *le dice* el ojo de la rana al cerebro de la rana?
48. Saltar por el espacio en la biblioteca de Babel
49. ¿Quién es el autor de *Spamlet*?
50. Ruido en el hotel virtual
51. Herb, Alice y Hal, el bebé
52. Memes
Resumen

VII. *Herramientas para pensar acerca de la conciencia*

53. Dos contraimágenes
54. La corazonada zombi
55. Zombis y zimbos
56. La maldición de la coliflor
57. Vim: ¿cuánto cuesta en «dinero de verdad»?
58. El triste caso del señor Clapgras
59. La baraja afinada
60. La habitación china
61. La caída del teleclón de Marte a la Tierra
62. El yo como centro de gravedad narrativa
63. Heterofenomenología
64. Mary, la científica del color: una muleta de razonamiento al descubierto
Resumen

VIII. *Herramientas para pensar acerca del libre albedrío*
65. Una neurocirujana verdaderamente nefanda
66. Un juguete determinista: el juego de la vida de Conway
67. Piedra, papel o tijeras
68. Dos loterías
69. Hechos históricos inertes
70. Maratón de ajedrez por computadora
71. Máxima responsabilidad
72. Sphexismo
73. Los niños del Brasil: otra muleta de razonamiento
Resumen

IX. *¿Cómo es ser un filósofo?*
74. Un trato faustiano
75. La filosofía como una autoantropología naïf
76. Verdades de orden superior del *ajetrez*
77. El diez por ciento que está bien

X. *Usa las herramientas. Esfuérzate más*

XI. *Lo que quedó fuera*

Apéndice. Soluciones a los problemas de la máquina de registro
Fuentes documentales

Bibliografía
Índice analítico

Transformar lo inconcebible en obvio, es decir, equipar al lector con los medios necesarios para pensar sobre cualquier tema, es el objetivo de Daniel Dennett al presentar en esta obra setenta y siete estrategias mentales a manera de una caja de herramientas del pensamiento: etiquetas, ejemplos, analogías, metáforas, "andamiajes", así como experimentos mentales bautizados por el autor como "bombas de intuición", son algunas de ellas. Se trata de herramientas no matemáticas, más bien informales o literarias, aquellas que la tradición científica suele desestimar pero que han sido la fuerza central de la filosofía durante siglos y que le han permitido a esta disciplina, mediante la creatividad y la imaginación, colocar el pensamiento humano en camino de resolver los grandes misterios.

www.fondodeculturaeconomica.com

Índice

Sumario 9
Prefacio 11
I. Introducción. ¿Qué es una bomba de intuición? 14
II. Doce herramientas de pensamiento generales 28
1. Cometer errores 30
2. «Por parodia de razonamiento»: el uso de la reductio ad absurdum 37
3. Las reglas de Rapoport 40
4. La ley de Sturgeon 42
5. La navaja de Ockham 44
6. La escoba de Ockham 46
7. Usar público lego como señuelo 48
8. Pensar afuera de la caja 50
9. Tres especies de gouldeo: masbienismo, apilamiento y el pasodoble de Gould 53
10. El operador «no cabe duda»: un blockeo mental 57
11. Preguntas retóricas 59
12. ¿Qué es una profulidad? 60
Resumen 62
III. Herramientas para pensar acerca del significado o contenido 66
13. Asesinato en la Plaza de Trafalgar 69
14. Un hermano mayor que vive en Cleveland 72
15. «Mi papá es doctor» 74
16. Imagen manifiesta e imagen científica 75
17. Psicología vulgar 78
18. La postura intencional 81
19. La distinción personal/subpersonal 88
20. Una cascada de homúnculos 92
21. El operador maomeno 96
22. Tejido maravilloso 98
23. Atrapado en el centro de operaciones del robot 101
IV. Intermedio sobre computadoras 109
24. Los siete secretos del poder de la computación revelados 111
25. Máquinas virtuales 131

26. Algoritmos 136
27. Automatizar el elevador 138
Resumen 142
V. Más herramientas sobre el significado 147
28. Algo con los pelirrojos 148
29. El tragacuartos ambulante, la Tierra Gemela y el robot gigante 151
30. Traducción radical y un crucigrama quineano 165
31. Motores semánticos y motores sintácticos 168
32. El Hombre del Pantano conoce a un tiburón vaca 170
33. Dos cajas negras 174
Resumen 185
VI. Herramientas para pensar acerca de la evolución 191
34. El ácido universal 193
35. La biblioteca de Mendel: Extremada y Evanescente 195
36. Los genes como palabras o como subrutinas 202
37. El árbol de la vida 204
38. Grúas y ganchos celestes: levantamiento en el Espacio de Diseño 205
39. Competencia sin comprensión 216
40. Razones de libre flotación 217
41. ¿Entienden los chapulines los números primos? 219
42. Cómo explicar los saltos de las gacelas 220
43. Cuidado con el Mamífero Primario 222
44. ¿Cuándo ocurre la especiación? 225
45. Hacedores de viudas, Eva Mitocondrial y coronaciones retroactivas 227
46. Ciclos 231
47. ¿Qué le dice el ojo de la rana al cerebro de la rana? 234
48. Saltar por el espacio en la biblioteca de Babel 236
49. ¿Quién es el autor de Spamlet? 238
50. Ruido en el hotel virtual 244
51. Herb, Alice y Hal, el bebé 247
52. Memes 250
Resumen 252
VII. Herramientas para pensar acerca de la conciencia 258
53. Dos contraimágenes 260
54. La corazonada zombi 261

55. Zombis y zimbos 265
56. La maldición de la coliflor 271
57. Vim: ¿cuánto cuesta en «dinero de verdad»? 274
58. El triste caso del señor Clapgras 277
59. La baraja afinada 283
60. La habitación china 289
61. La caída del teleclón de Marte a la Tierra 297
62. El yo como centro de gravedad narrativa 299
63. Heterofenomenología 305
64. Mary, la científica del color: una muleta de razonamiento al descubierto 310
Resumen 314
VIII. Herramientas para pensar acerca del libre albedrío 320
65. Una neurocirujana verdaderamente nefanda 323
66. Un juguete determinista: el juego de la vida de Conway 325
67. Piedra, papel o tijeras 334
68. Dos loterías 338
69. Hechos históricos inertes 340
70. Maratón de ajedrez por computadora 345
71. Máxima responsabilidad 352
72. Sphexismo 355
73. Los niños del Brasil: otra muleta de razonamiento 359
Resumen 363
IX. ¿Cómo es ser un filósofo? 368
74. Un trato faustiano 370
75. La filosofía como una autoantropología naïf 373
76. Verdades de orden superior del ajetrez 376
77. El diez por ciento que está bien 381
X. Usa las herramientas. Esfuérzate más 386
XI. Lo que quedó fuera 389
Apéndice. Soluciones a los problemas de la máquina de registro 394
Fuentes documentales 406
Bibliografía 420
Índice analítico 436