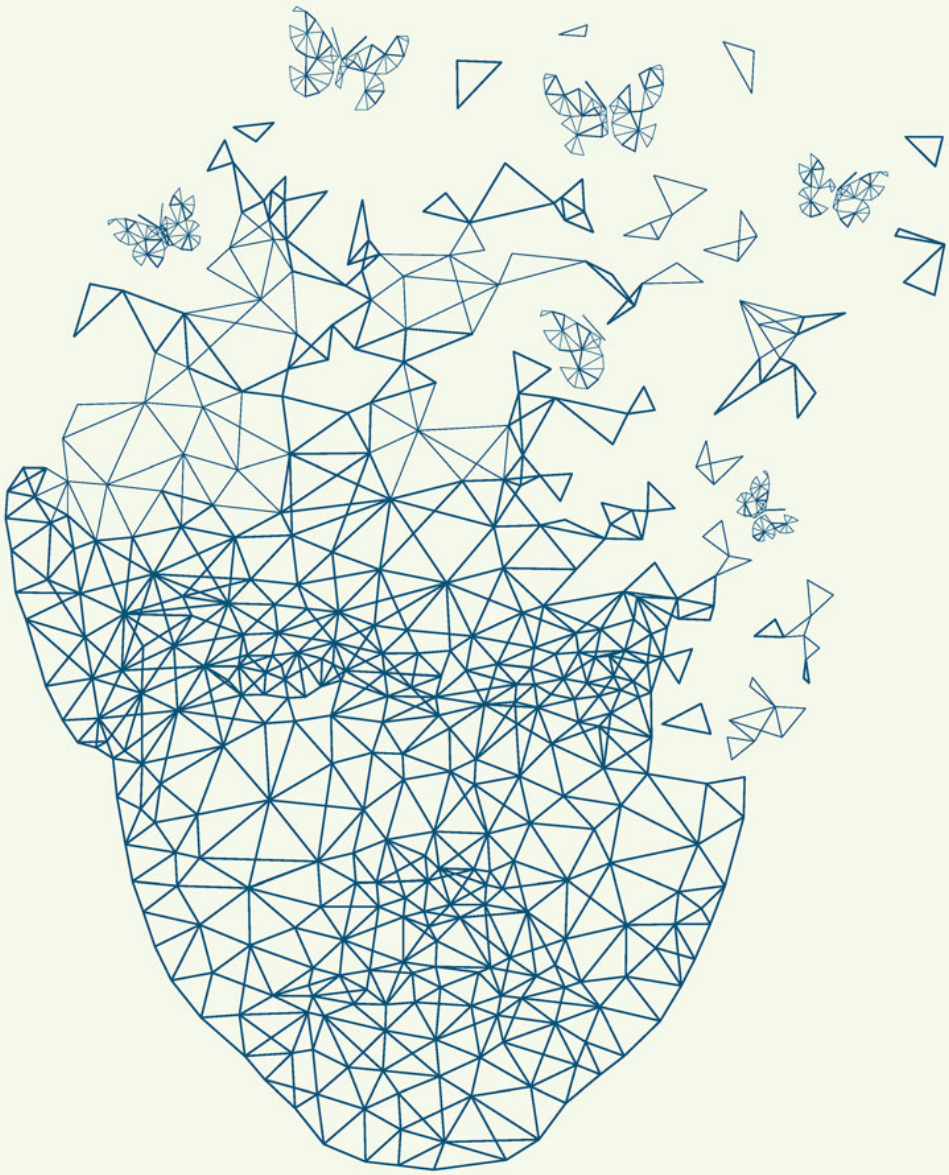
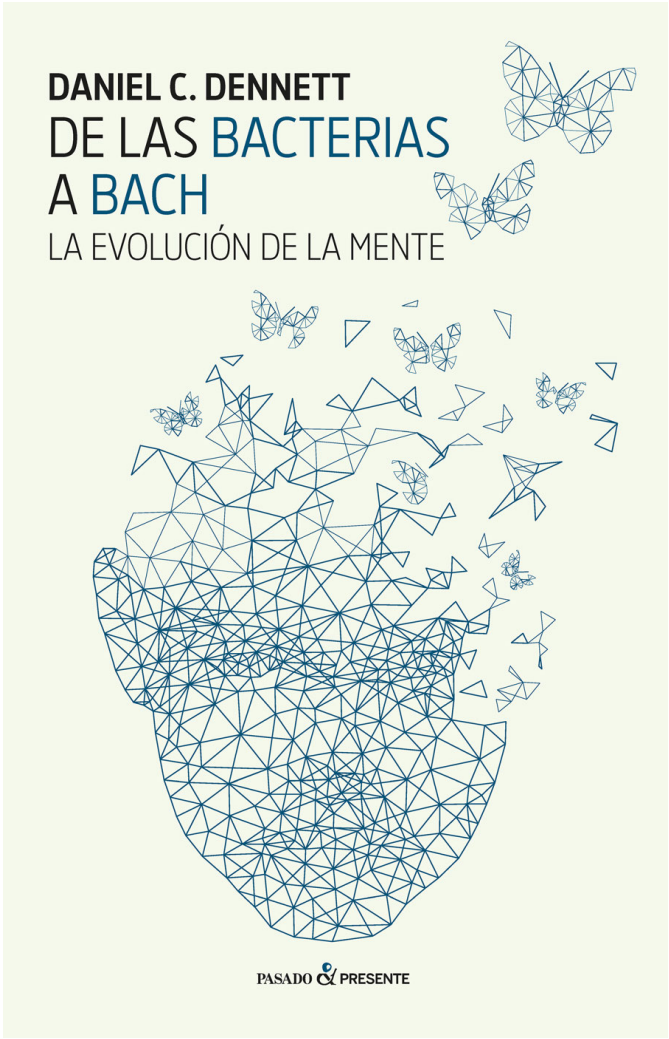
DANIEL C. DENNETT

DE LAS BACTERIAS A BACH

LA EVOLUCIÓN DE LA MENTE



PASADO & PRESENTE


DANIEL C. DENNETT
DE LAS BACTERIAS
A BACH
LA EVOLUCIÓN DE LA MENTE
PASADO & PRESENTE

DE LAS BACTERIAS A BACH

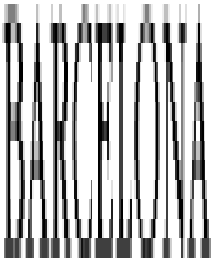
DANIEL C. DENNETT

DE LAS BACTERIAS A BACH

La evolución de la mente

Traducción de

MARC FIGUERAS



A Brandon, Samuel, Abigail y Aria

PREFACIO

Empecé a pensar seriamente acerca de la evolución de la mente humana durante mi doctorado en filosofía en Oxford en 1963, una época en la que no sabía casi nada sobre la evolución ni sobre la mente humana. En aquel entonces no se esperaba que los filósofos fueran expertos en temas científicos y, como resultado, incluso los más eminentes eran en buena parte ignorantes sobre trabajos en psicología, neuroanatomía y neurofisiología (los términos ciencia cognitiva y neurociencia no se acuñarían hasta una década más tarde). El joven campo denominado inteligencia artificial por John McCarthy en 1956 atraía una atención considerable, pero muy pocos filósofos habían siquiera tocado una computadora, esas máquinas que zumbaban misteriosamente en prisiones de aire acondicionado vigiladas por técnicos. Así pues, era el momento ideal para que un mero aficionado sin formación como yo aprendiera algo sobre todos estos temas. Un filósofo que hacía preguntas interesantes sobre lo que estaban haciendo (en lugar de decirles por qué sus proyectos eran imposibles en principio) parecía una novedad tan refrescante que un admirable grupo de expertos me acogió en su seno, me enseñó de manera informal y me aconsejó a quién tomarme en serio y qué textos leer, a la vez que perdonaba mis ingenuos malentendidos con más benevolencia que si hubiera sido uno de sus colegas o estudiantes.

Hoy en día hay docenas, o centenares, de jóvenes filósofos con una sólida formación interdisciplinaria en ciencia cognitiva, neurociencia e informática, con un nivel mucho más alto que el que yo tenía. Algunos son estudiantes míos, o incluso estudiantes de mis estudiantes, pero otros filósofos de mi generación se lanzaron al vacío (a menudo con más formación que yo) y ahora tienen sus equipos de estudiantes, haciendo progresos en los campos más punteros, ya sea como filósofos interdisciplinarios o como científicos con formación filosófica con sus propios grupos de investigación. Son profesionales, mientras que yo sigo siendo un aficionado, aunque ahora ya puedo decir que soy un aficionado con conocimientos profundos, al que invitan a impartir conferencias, a participar en seminarios y a visitar laboratorios por todo

el mundo, donde continúo mi formación y me divierto más de lo que nunca hubiera imaginado que podía con una vida académica.

Entre otras cosas, considero que este libro es mi intento de agradecimiento como pago por mi matrícula en toda la formación que he recibido. Aquí presento lo que creo que he aprendido; y buena parte de ello es aún hipotético, filosófico, arriesgado. Mi propuesta es que se trata del esquema, el esqueleto de la mejor teoría científica que tenemos hoy para explicar cómo surgieron nuestras mentes, cómo consiguen los cerebros obrar todas sus maravillas y, especialmente, cómo hay que pensar acerca de las mentes y los cerebros sin caer en seductoras trampas filosóficas. Es una afirmación controvertida, sin duda, y ya estoy deseando enfrentarme a las reacciones tanto de filósofos como de científicos, así como de los aficionados, que suelen exponer los comentarios más perspicaces.

Muchas personas me han ayudado al escribir mis libros, pero aquí me centraré en las que me han ayudado específicamente con las ideas presentadas en este, quienes, faltaría más, no son en absoluto responsables de los errores que no han podido evitar. Así, quiero dar las gracias a los participantes en el grupo de trabajo que organicé en el Santa Fe Institute sobre evolución cultural en mayo de 2014: Sue Blackmore, Rob Boyd, Nicolas Claidière, Joe Henrich, Olivier Morin, Pete Richerson, Peter Godfrey-Smith, Dan Sperber y Kim Sterelny y otros del SFI, sobre todo Chris Wood, Tanmoy Bhattacharya, David Wolpert, Cris Moore, Murray Gell-Mann y David Krakauer. También quiero expresar mi gratitud a Louis Godbout de la Sybilla Hesse Foundation, por respaldar el seminario.

También tenemos a mis estudiantes de Tufts y a los asistentes que participaron en un seminario en la primavera de 2015, cuando estaba redactando los primeros borradores de muchos capítulos del libro: Alicia Armijo, Edward Beuchert, David Blass, Michael Dale, Yufei Du, Brendan Fleig-Goldstein, Laura Friedman, Elyssa Harris, Justis Koon, Runeko Lovell, Robert Mathai, Jonathan Moore, Savannah Pearlman, Nikolai Renedo, Tomas Ryan, Hao Wan, Chip Williams, Oliver Yang y Daniel Cloud, que asistió al seminario para comentar su nuevo libro. Y Joan Vergés-Gifra, Eric Schliesser, Pepa Toribio, Mario Santos Sousa y el resto del animado grupo que se reunió en la Universidad de Gerona,

en la que pasé una intensa semana de mayo como profesor invitado para la cátedra Ferrater Mora de pensamiento contemporáneo. Otro banco de pruebas vino de la mano de Anthony Grayling y el profesorado y los estudiantes que reunió en el New College of the Humanities de Londres, donde he estado probando versiones de estas ideas durante los últimos cuatro años.

Otros nombres que se han peleado con mis borradores, me han hecho cambiar de ideas, han señalado mis errores y me han alentado a buscar mayor claridad son Sue Stafford, Murray Smith, Paul Oppenheim, Dale Peterson, Felipe de Brigard, Bryce Huebner, Enoch Lambert, Amber Ross, Justin Junge, Rosa Cao, Charles Rathkopf, Ronald Planer, Gill Shen, Dillon Bowen y Shawn Simpson. También he recibido buenos consejos de Steve Pinker, Ray Jackendoff, David Haig, Nick Humphrey, Paul Seabright, Matt Ridley, Michael Levin, Jody Azzouni, Maarten Boudry, Krys Dolega, Frances Arnold y John Sullivan.

Como con mi libro anterior, Bombas de intuición y otras herramientas de pensamiento, los editores Drake McFeely y Brendan Curry de Norton me han impulsado a clarificar, simplificar, reducir, ampliar, explicar y también suprimir, convirtiendo la obra final en una experiencia de lectura mucho más unificada y eficaz de lo que hubiera sido sin su ayuda. Como siempre, John Brockman y Katinka Matson han sido unos agentes literarios perfectos, aconsejando, alentando y divirtiendo (y, por supuesto, vendiendo) al autor, tanto en casa como fuera. Teresa Salvato, coordinadora de programas en el Center for Cognitive Studies, ha gestionado toda la logística de mi vida académica durante años, dándome miles de horas para escribir e investigar, y ha tenido un papel clave en este libro, buscando libros y artículos en las bibliotecas y organizando las referencias.

Para finalizar, quiero dar las gracias a mi esposa Susan, que ha sido mi sostén, consejera, crítica y mejor amiga durante más de medio siglo y siempre ha mantenido la cantidad justa de carbón en la caldera para que la cocción siguiera a fuego lento, con todos los altibajos del camino; merece todos los elogios posibles por su contribución a nuestra empresa conjunta.

DANIEL DENNETT

North Andover, Massachusetts

28 de marzo de 2016

PRIMERA PARTE

EL MUNDO, PATAS ARRIBA

1

INTRODUCCIÓN

BIENVENIDOS A LA SELVA

¿Cómo es que hay mentes? ¿Y cómo es posible que estas mentes hagan esta pregunta y la respondan? La respuesta breve es que las mentes han evolucionado y han creado recursos de pensamiento que, a la larga, les han permitido saber cómo han evolucionado las mentes e, incluso, les han permitido saber cómo estos recursos les han permitido saber qué son las mentes. ¿De qué recursos estamos hablando? El más simple, del cual dependen todos los demás en formas diversas, es el lenguaje hablado, seguido de la lectura, la escritura y la aritmética, para pasar a la navegación y la cartografía, el aprendizaje y todos los diversos artefactos que hemos inventado para obtener y manipular información: la brújula, el telescopio, el microscopio, la cámara fotográfica, el ordenador, Internet y tantos otros. Estos artefactos inundan nuestras vidas de ciencia y tecnología y nos permiten saber muchas cosas desconocidas para otras especies; nosotros sabemos que hay bacterias, pero los perros no, ni los delfines, ni los chimpancés; ni tan solo las bacterias saben que hay bacterias. Nuestras mentes son diferentes; se necesitan recursos de pensamiento para entender qué son las bacterias, y resulta que nosotros somos la única especie (hasta el momento) que dispone de un intrincado conjunto de tales recursos.

Esta es la respuesta breve, y reducida a su mínima expresión no debería resultar polémica, pero tras los detalles acechan diversas implicaciones imprevistas, quizá impactantes, que todavía no se aprecian ni se comprenden correctamente. A partir de la anodina suposición inicial de que las personas somos objetos físicos, sometidos a las leyes de la física, un camino tortuoso, que atraviesa una selva de ciencia y filosofía, nos lleva hasta la comprensión de nuestras mentes conscientes. El camino está

repleto de dificultades, empíricas y conceptuales, y muchos expertos discrepan vehementemente acerca del modo en que deben abordarse tales problemas. Durante más de cincuenta años he intentado atravesar todos estos matorrales y cenagales y he hallado una ruta que nos conduce a una descripción satisfactoria y gratificante de cómo se produce la «magia» de nuestras mentes sin ningún tipo de magia; pero no es una ruta fácil ni directa. No es el único camino que se nos presenta, pero, tal como espero mostrar, hasta la fecha es el mejor y el más prometedor. Para emprender el viaje es necesario abandonar algunas preciadas intuiciones, pero creo que, por fin, he encontrado la manera de hacer que el acto de descartar estas «verdades evidentes» no solo sea soportable sino también agradable: en cierto modo, vuelve nuestra mente del revés y nos ofrece unos puntos de vista sorprendentes acerca de qué es lo que está pasando. Ahora bien, deberemos abandonar algunas ideas apreciadas por muchos.

Diversos eminentes pensadores discrepan de las propuestas que he ido planteando a lo largo de los años, y supongo que algunos seguirán encontrando mis nuevas incursiones tan intolerables como las anteriores, pero actualmente empiezo a encontrar una agradable compañía a lo largo de mi camino, nuevos respaldos para los hitos que propongo y nuevos temas para incitar a las diversas y extrañas inversiones del razonamiento que invitaré a hacer. Algunas serán familiares para los que hayan leído mis obras anteriores, pero he reparado, reforzado y rediseñado estas ideas para soportar un peso mucho más grande que antes. A primera vista, las nuevas son tan contrarias al sentido común como las anteriores, y por mi experiencia de muchos años mirando de persuadir a la gente paso a paso (y fracasando en el intento), intentar apreciarlas sin seguir mi sinuoso camino puede ser un esfuerzo vano. A continuación, os presento una lista de algunos de los obstáculos (para el pensamiento acomodado) que encontraréis en mi camino, pero que no espero que «asumáis» en un primer encuentro:

La extraña inversión del razonamiento de Darwin

Razones sin razonadores

Competencia sin comprensión

La extraña inversión del razonamiento de Turing

La información como diseño que merece la pena robar

El darwinismo del darwinismo

Neuronas salvajes

Palabras que luchan por reproducirse

La evolución de la evolución de la cultura

La extraña inversión del razonamiento de Hume

La conciencia como una ilusión de los usuarios

La era del posdiseño inteligente

«¿Información como diseño que merece la pena robar? ¿Acaso no has oído hablar de la teoría matemática de la información de Shannon?» «¿Neuronas salvajes? ¿En contraposición a qué? ¿A neuronas domesticadas?» «¿Estás de broma? ¿La conciencia como una ilusión? ¿En serio?»

Si no fuera por los cada vez más numerosos teóricos, científicos y filósofos bien documentados, que están de acuerdo con buena parte de mis puntos de vista y que han contribuido a ellos, sin duda habría perdido todo mi coraje y habría concluido que soy yo quien está fatalmente desconcertado. También es posible, ciertamente, que nuestra audaz comunidad no sea más que un grupo de entusiastas que se están engañando mutuamente; pero antes de emitir un veredicto, veamos cómo van las cosas.

Sé perfectamente lo fácil y tentador que es ignorar estas curiosas ideas o descartarlas sin prestarles más atención la primera vez que uno se encuentra con ellas, porque es algo que yo mismo he hecho a menudo. Me recuerdan esos enigmas que, retrospectivamente, tienen una solución obvia que al inicio pasas por alto con un juicio instantáneo del tipo «no puede ser eso» o ni tan solo la consideras de tan poco prometedora que parece.[1] Para alguien que a menudo ha acusado a otros de confundir fracasos de la imaginación con intuiciones perspicaces, es embarazoso reconocer mis propios lapsus en este tema, pero tras haber tropezado con

nuevas formas de plantear los problemas, estoy ansioso por llevar mis soluciones recién halladas a los grandes enigmas acerca de la mente. Estas doce ideas, y el trasfondo necesario para hacerlas más apetitosas, se presentarán, aproximadamente, en el orden en que las acabo de mostrar. Aproximadamente, porque he visto que algunas se resisten a una defensa clara: no las puedes entender hasta que ves lo que te pueden aportar, pero no las puedes usar hasta que las entiendes, de modo que hay que empezar con una presentación parcial que esboce la idea y luego volver a ella tras verla en acción, hasta tomar plena conciencia de la misma.

El razonamiento del libro está formado por tres arduos ejercicios de imaginación:

poner nuestro mundo patas arriba, siguiendo a Darwin y Turing;

luego, hacer evolucionar la evolución en diseño inteligente;

finalmente, volver nuestras mentes del revés.

La base debe cimentarse cuidadosamente durante los cinco primeros capítulos, si queremos que soporte nuestras suposiciones para el segundo hito. Los siguientes ocho capítulos ahondan en los detalles empíricos de la evolución de la mente y el lenguaje, tal como se ven desde nuestro punto de vista invertido. Esto nos permitirá enmarcar nuevas preguntas y esbozar nuevas respuestas, que prepararán el terreno para la inversión más difícil de todas: ver cómo se observa la conciencia desde el nuevo punto de vista.

Es un camino exigente, pero en algunos trechos repasaré temas familiares, para asegurarme de que todo el mundo sigue al mismo ritmo. Los lectores que conozcan estos temas mejor que yo pueden saltárselos, si quieren, o pueden aprovechar mi tratamiento de estos para valorar cuánto pueden confiar en mí acerca de los temas sobre los que no saben tanto.

Emprendamos la marcha.

NUESTRO VIAJE A VISTA DE PÁJARO

La vida ha evolucionado en este planeta durante casi cuatro mil millones de años. Los primeros dos mil millones, aproximadamente, estuvieron dedicados a optimizar la maquinaria básica para la sustentación, la obtención de energía y la reproducción; los únicos seres vivos eran entidades unicelulares comparativamente simples, los procariotas: las bacterias, y sus parientes, las arqueas. Luego sucedió una cosa extraordinaria; dos procariotas diferentes chocaron, cada uno con sus capacidades y hábitos debidos a miles de millones de años de evolución independiente. Con toda seguridad, choques de este tipo se habían producido incontables veces, pero en una de tales ocasiones (como mínimo), una célula envolvió a la otra y, en lugar de destruirla y usar sus partes como combustible o material de construcción (es decir, comérsela), la dejó vivir en su interior y, por pura suerte, resultó que estaba más adaptada (más competente en algunos aspectos relevantes) que cuando iba a solas, sin ninguna carga.

Probablemente, este fue el primer caso satisfactorio de transferencia de tecnología, la situación en que dos conjuntos de capacidades, perfeccionados a lo largo de eones de I+D (investigación y desarrollo) independiente, se unen para formar algo más grande y mejor. A diario aparecen noticias sobre Google, Amazon o General Motors devorando alguna pequeña empresa emergente para hacerse con sus innovaciones tecnológicas y sus perspicaces avances en I+D, más fáciles de conseguir en unas abarrotadas oficinas que en enormes empresas; sin embargo, el uso original de esta táctica dio a la evolución su primer gran impulso. Las uniones al azar no siempre funcionan de esta manera; de hecho, casi nunca lo hacen, pero la evolución es un proceso que se basa en amplificar cosas que no suceden casi nunca. Por ejemplo, las mutaciones en el ADN no se producen casi nunca (ni una sola vez de cada mil millones de copias), pero la evolución depende de ellas. Es más, la gran mayoría de mutaciones son perjudiciales o neutras; una mutación casualmente «buena» es algo que casi nunca sucede. Sin embargo, la evolución se basa en estos acontecimientos, los más excepcionales de entre los menos frecuentes.

La especiación, el proceso por el cual se crea una nueva especie cuando algunos miembros de una población quedan aislados de su grupo original

y vagan por el espacio genético para formar un nuevo patrimonio genético, es un suceso muy poco frecuente, pero los millones o miles de millones de especies que han existido en este planeta empezaron a partir de un episodio de especiación. Cada nacimiento de cada linaje es un posible inicio de un acontecimiento de especiación, pero esta no sucede casi nunca, ni tan solo una vez por cada millón de nacimientos.

En el caso que estamos considerando, la infrecuente mejora que apareció a partir del choque fortuito de una bacteria y una arquea tuvo un resultado que cambió la vida. Al estar mejor adaptado, este dúo se reprodujo con más eficacia que la competencia, y cada vez que se dividía en dos (la forma que tienen las bacterias de reproducirse) las dos células hijas incluían también la descendencia del huésped original. De este modo, sus destinos quedaron unidos (se produjo una simbiosis) gracias a uno de los sucesos más fructíferos de la historia de la evolución. Se trata de una endosimbiosis porque uno de los miembros está, literalmente, dentro del otro; a diferencia de la ectosimbiosis externa del pez payaso y las anémonas o de los hongos y las algas en los líquenes. De este modo nació la célula eucariota que, al tener más componentes, era más versátil que sus antecesoras, las células procariotas simples, como las bacterias.[2] A lo largo del tiempo, estos eucariotas se hicieron más grandes, más complejos y más competentes, en definitiva, mejores (el prefijo eu- en eucariota es el mismo que en eugenesia, eufonía o eutanasia, que significa ‘bueno’). Los eucariotas fueron el ingrediente clave para que se pudiera formar la vida multicelular en todas sus variedades. En una primera aproximación, cualquier ser vivo lo bastante grande como para ser observado a simple vista es un eucariota multicelular. Nosotros somos eucariotas, igual que lo son los tiburones, los pájaros, los árboles, los champiñones, los insectos, los gusanos y todos los demás animales y plantas, descendientes directos, todos ellos, de la célula eucariota original.

Esta revolución eucariótica sentó las bases de otra gran transición, la explosión del Cámbrico de hace más de 500 millones de años, en la que se produjo la aparición «repentina» de una gran cantidad de nuevas formas de vida. Más adelante llegó lo que yo denomino la explosión de MacCready, a partir del gran ingeniero y visionario Paul MacCready (creador del Gossamer Albatross, entre otras maravillas ambientalistas). A diferencia de la diversificación cámbrica, que se produjo a lo largo de varios millones de años, hace 530 millones de años (Gould 1989), la explosión de MacCready se dio en solo 10.000 años, o 500 generaciones humanas. Según los cálculos de MacCready (1999), en los albores de la

agricultura humana, hace 10.000 años, la población humana mundial, más el ganado y los animales domésticos, solo representaba alrededor del 0,1% de la biomasa de vertebrados terrestres (no estamos incluyendo ni los insectos ni otros invertebrados ni los animales marinos); actualmente, según sus estimaciones, esta cifra es del 98% (la mayoría de la cual corresponde al ganado). Vale la pena citar sus pensamientos acerca de esta sorprendente expansión:

A lo largo de miles de millones de años, en una esfera única, el azar ha pintado una fina capa de vida; compleja, improbable, maravillosa y frágil. De golpe, los humanos […] hemos aumentado nuestra población, nuestra tecnología y nuestra inteligencia hasta llegar a una posición de tremendo poder: actualmente somos nosotros quienes manejamos la brocha. (1999, p. 19)

En nuestro planeta ha habido otros cambios comparativamente rápidos, extinciones masivas como la gran extinción del Cretácico-Paleógeno hace unos 66 millones de años, que condenó a los dinosaurios; pero ciertamente la explosión de MacCready es uno de los cambios biológicos más rápidos que se hayan producido jamás en la Tierra. Aún está en marcha, y acelerando. Podemos salvar el planeta o extinguir toda la vida que hay en él, algo que ninguna otra especie puede ni tan solo concebir. Parece evidente que el orden en que MacCready enumera sus tres factores (población, tecnología e inteligencia) debería invertirse: primero, la inteligencia humana creó la tecnología (incluida la agricultura) que, luego, permitió el auge de población; sin embargo, como hemos visto, la evolución suele ser una intrincada red llena de recovecos coevolutivos: nuestra denominada inteligencia natural depende, de un modo a menudo imprevisto, tanto de nuestra tecnología como de nuestra población.

Las mentes humanas son sorprendentemente diferentes de las mentes de otras especies, son mucho más potentes y versátiles. La respuesta larga a cómo es que tenemos unas mentes tan notables empieza a esclarecerse. En una famosa cita, el biólogo británico D'Arcy Thompson (1917) dijo que «todo es como es porque se produjo de esa manera». Muchos de los enigmas (o misterios, o paradojas) de la conciencia humana se esfuman en cuanto preguntas cómo pueden haber surgido… ¡e intentas responder a

la pregunta! Remarco esto último porque hay personas que se maravillan con la pregunta y luego «responden» afirmando «es un misterio inescrutable» o «Dios lo hizo así». Por supuesto, podría ser que estuvieran en lo cierto, pero teniendo en cuenta la extraordinaria abundancia de recursos de pensamiento que tenemos a nuestra disposición y que apenas se han usado todavía, es una capitulación sorprendente y prematura. Puede que no sea una actitud de derrotismo, sino una actitud defensiva. A algunas personas les gustaría convencer a los curiosos de que sacaran las manos de sus queridos misterios, sin darse cuenta de que un misterio resuelto es mucho más cautivador que las ignorantes fantasías a las que sustituye. Otras personas han examinado en profundidad las explicaciones científicas y no están de acuerdo con ellas; en su opinión, los antiguos mitos de carros ardientes, dioses enfrentados, mundos que surgen de huevos de serpiente, encantamientos malignos y jardines fabulosos son más agradables y dignos de atención que cualquier argumento científico predictivo. Está claro que es imposible complacer a todo el mundo.

Esta pasión por el misterio no es más que uno de los poderosos inhibidores de la imaginación que bloquean nuestra ruta cuando intentamos responder a la pregunta de cómo es que hay mentes. Como ya he señalado, nuestro camino tendrá que volver atrás varias veces, volviendo a preguntas que habremos dejado aplazadas porque no se podían responder hasta disponer de información que no podemos obtener hasta tener los recursos necesarios, recursos con los que, a su vez, no podemos contar hasta saber de dónde vienen; un proceso cíclico que, paso a paso, va completando los detalles de un esbozo que no resultará convincente hasta que lleguemos a un mirador desde donde podamos volver la vista atrás y ver cómo encajan todos los elementos.

El libro de Douglas Hofstadter Yo soy un extraño bucle (2007) describe una mente que se forma a sí misma en bucles de procesamiento que van dando vueltas, girando y sustentándose a sí mismos, creando frondosas reacciones a las reflexiones, a los recuerdos, a las reevaluaciones, y generando así nuevas estructuras: ideas, fantasías, teorías y, sin duda, recursos de pensamiento para crear aún más estructuras. Os recomiendo que lo leáis, conducirá a vuestra imaginación por una montaña rusa y descubriréis muchas verdades sorprendentes. En el libro que tenéis ahora en las manos, mi propósito es el bucle extraño más grande (formado por procesos formados por procesos) que ha generado mentes como la de Hofstadter (y la de Bach, y la de Darwin) a partir de nada más que

moléculas (formadas por átomos formados por…). Como se trata de una tarea cíclica, tendremos que empezar en algún lugar a medio camino e ir y volver varias veces. Además, la tarea se ve dificultada por un rasgo que no comparte con otras investigaciones científicas de los procesos (en cosmología, geología, biología o historia, por ejemplo): a las personas les afectan tanto las posibles respuestas a esta cuestión que cuesta mucho que las consideren de manera objetiva.

Por ejemplo, es posible que algunos lectores ya estén sacudiendo la cabeza por culpa de una afirmación que acabo de hacer: «las mentes humanas son sorprendentemente diferentes de las mentes de otras especies, son mucho más potentes y versátiles». ¿Realmente puedo tener estos prejuicios? ¿Acaso soy un «chovinista de la especie», un «especifista», que realmente cree que la mente humana es mucho más extraordinaria que la mente de los delfines, de los elefantes, de los cuervos, de los bonobos y de otras tantas especies cuyos talentos cognitivos hemos descubierto en los últimos años, y ante los cuales nos hemos maravillado? ¿No es esto un ejemplo descarado de la falacia del «excepcionalismo humano»? Quizá algunos lectores estén ya a punto de lanzar el libro por la ventana; otros puede que estén solo un poco inquietos por mi desliz políticamente incorrecto. Es gracioso (para mí, al menos) que el excepcionalismo humano provoque el mismo grado de indignación en sentidos opuestos. Algunos científicos y muchos defensores de los derechos de los animales lo condenan como uno de los peores pecados intelectuales, un vestigio despreciable y científicamente ignorante de los viejos tiempos en que la gente pensaba que todos los animales habían sido colocados en este planeta para nuestro provecho y diversión. Remarcan que nuestros cerebros están hechos de las mismas neuronas que los cerebros de un pájaro y que algunos son tan grandes como los nuestros (y tan inteligentes, a su propia manera, específica de su especie). Cuanto más se estudian las situaciones y comportamientos reales de los animales en estado salvaje, más se puede apreciar su genialidad. Otros pensadores, especialmente en los campos del arte, las humanidades y las ciencias sociales, consideran que negar el excepcionalismo humano es una postura miope y doctrinaria, cientificismo de la peor especie; ¡por supuesto que nuestra mente es órdenes de magnitud más potente que la mente del animal más inteligente! Ningún animal crea arte, escribe poesía, elabora teorías científicas, construye naves espaciales, surca los océanos, ni tan siquiera domina el fuego. Ante esto se esgrime la réplica: ¿qué hay de los elaborados emparrados que construyen los pergoleros, de las sutilezas

políticas de los chimpancés, de la pericia navegadora de las ballenas, de los elefantes y las aves migratorias, del virtuoso canto de los ruiseñores, del lenguaje de los cercopitecos o de las abejas? A su vez, esto provoca la respuesta de que estas maravillas del mundo animal son logros irrisorios en comparación con el genio de los artistas, ingenieros y científicos humanos. Hace algunos años,[3] acuñé los términos romántico y aguafiestas para referirme a los dos bandos de este intenso duelo sobre las mentes de los animales. Uno de mis recuerdos preferidos de esta reacción bipolar a las afirmaciones acerca de la inteligencia de los animales se produjo en un seminario internacional sobre este tema, cuando un famoso investigador consiguió representar tanto el papel de romántico como el de aguafiestas, y ambos con la misma pasión: «¡Ja! ¿Piensas que los insectos son estúpidos? Te voy a demostrar lo inteligentes que son. Fíjate en estos resultados…» para luego, el mismo día, afirmar: «¿Crees que las abejas son muy inteligentes? Deja que te demuestre cuán estúpidas son, en realidad; ¡si solo son pequeños robots mecánicos!»

¡Haya paz! Veremos que ambos bandos tienen razón en algunas cosas y están equivocados en otras. No somos los genios divinos que a veces nos creemos, pero los animales tampoco son tan inteligentes y, sin embargo, tanto los humanos como el resto de animales están maravillosamente capacitados para enfrentarse con brillantez a los retos que les plantea un mundo siempre complejo, si no cruel. Y nuestras mentes humanas son excepcionalmente potentes de maneras que solo ahora empezamos a comprender, al ver cómo han llegado a ser de la manera que son.

¿Por qué esto nos preocupa tanto? Esta es una de las muchas cuestiones pendientes que merece una respuesta, pero no ahora mismo, aunque sí daré un resumen muy breve: aunque los procesos que dieron lugar a esta preocupación se remontan a miles de años (y en ciertos aspectos, a millones o miles de millones de años), solo se convirtieron en un tema, en un objeto sobre el que pensar y preocuparse, con el nacimiento de la ciencia moderna en el siglo XVII. Por eso es justo aquí cuando saltaré a la palestra y daré inicio a esta versión de la historia.

LA HERIDA CARTESIANA

Sì, abbiamo un anima. Ma è fatta di tanti piccoli robot!

(Sí, tenemos un alma, ¡pero está hecha de muchos pequeños robots!)

Titular de una entrevista que me hizo Giulio Giorello

para el Corriere della Sera, Milán, 1997

René Descartes, el científico y filósofo francés del siglo XVII, estaba muy impresionado con su propia mente, y no le faltaba razón. La llamó res cogitans, 'cosa pensante', y reflexionando sobre ella, le pareció algo de unas capacidades milagrosas. Si ha habido alguien con el derecho de maravillarse de su propia mente, Descartes sin duda lo era. Fue uno de los científicos más grandes de todos los tiempos, con trabajos importantes en matemáticas, óptica, física y fisiología. También fue el inventor de uno de los recursos de pensamiento más valiosos de la historia: el sistema de «coordenadas cartesianas» que permite pasar del álgebra a la geometría y viceversa; este invento allanó el camino para el desarrollo del análisis matemático y con él podemos representar gráficamente casi cualquier cosa que queramos investigar, desde el crecimiento de los cerdos hormigueros hasta los contratos de futuros del zinc. Descartes propuso la primera teoría del todo, una gran teoría unificada prototípica que publicó bajo el poco modesto título de Le Monde. Pretendía explicarlo todo, desde las órbitas de los planetas y la naturaleza de la luz a las mareas, desde los volcanes a los imanes, por qué el agua se agrupa en gotas esféricas, cómo saltan chispas del pedernal y muchísimas otras cosas. Su teoría era errónea casi en su totalidad, pero estaba muy bien conjuntada e incluso desde nuestro punto de vista actual tiene un aspecto extrañamente plausible. Le tocó a Isaac Newton elaborar una física mejor que, publicada en sus famosos Principia, refutaba explícitamente la teoría de Descartes.

Descartes no pensaba que solamente su mente fuera maravillosa; todas las mentes humanas normales eran extraordinarias, capaces de logros que ningún animal podía emular, logros que estaban fuera del alcance de cualquier mecanismo imaginable, por muy complejo y sofisticado que fuera. De ello, llegó a la conclusión de que las mentes como la suya (y las

nuestras) no eran entes materiales, como los pulmones o el cerebro, sino entes formados por un segundo tipo de sustancia que no estaba constreñida a obedecer las leyes de la física. Así formuló el punto de vista conocido como dualismo o, también, dualismo cartesiano. La idea de que la mente no es materia y de que la materia no puede formar la mente no la inventó Descartes; desde hacía miles de años, a los pensadores les había parecido evidente que nuestras mentes no son como el «mobiliario» del mundo externo. La doctrina de que cada uno de nosotros tiene un alma inmaterial (e inmortal) que reside en el cuerpo material y lo controla se ha ido transmitiendo como un conocimiento compartido, gracias a los preceptos de la Iglesia. Aun así, fue Descartes quien perfiló esta suposición básica en una «teoría» clara, según la cual la mente inmaterial, la «cosa pensante» consciente que conocemos íntimamente mediante la introspección, está de algún modo en comunicación con el cerebro material, que aporta toda la información, pero en ningún caso la comprensión ni las experiencias.

El problema del dualismo, ya desde Descartes, es que nunca nadie ha podido ofrecer una descripción convincente de cómo se produce esta supuesta interacción entre la mente y el cuerpo sin violar las leyes de la física. Las candidatas actuales nos dan a elegir entre una revolución en la ciencia tan radical que es imposible de describir (lo que resulta conveniente, porque los críticos están al acecho, listos para abalanzarse sobre ella) o una declaración de que algunas cosas, al fin y al cabo, son simplemente misterios, más allá de la comprensión humana (lo que también es muy conveniente si no tienes ninguna idea y quieres dejar el tema rápidamente). Pero incluso si el dualismo, tal como remarqué hace años, tiende a ser considerado como un precipicio por el borde del cual puedes empujar a tus oponentes, a los que se quedan arriba les queda mucho trabajo por hacer si quieren construir una teoría que no sea una especie de dualismo disfrazado. La misteriosa relación entre la mente y la materia ha sido un campo de batalla de científicos y filósofos desde el siglo XVII.

Francis Crick, codescubridor de la estructura del ADN, ha sido otro de los grandes científicos de la historia, fallecido en 2003; su última gran obra se titula La búsqueda científica del alma (1994), en la cual argumenta que el dualismo es falso; es decir, que la mente es el cerebro, un órgano material sin misteriosas propiedades adicionales imposibles de encontrar en otros seres vivos. Crick no fue el primero en plantear esta negación del dualismo; ha sido la opinión predominante (aunque no unánime) entre

científicos y filósofos durante casi un siglo. El título original del libro, en inglés, es The Astonishing Hypothesis (‘la hipótesis asombrosa’) y, de hecho, muchos de los que nos dedicamos a este campo no estuvimos de acuerdo con tal título: ¡no había nada de asombroso en esta hipótesis! ¡era nuestra hipótesis de trabajo desde hacía décadas! Lo asombroso sería negarla, algo así como afirmar que el oro no está formado por átomos o que la ley de la gravedad no se cumple en Marte. ¿Por qué deberíamos esperar que la conciencia escindiera el universo de una forma tan tajante, cuando incluso la vida y la reproducción se pueden explicar en términos físico-químicos? Pero Crick no escribió su libro para los científicos y los filósofos, y sabía que, entre los profanos, el dualismo tiene aún un atractivo abrumador: no solo parece evidente que los pensamientos y experiencias privadas de cada persona se tienen que llevar a cabo en algún medio diferente de aquel por el que se propagan los impulsos nerviosos que los científicos han detectado en sus cerebros, sino que la perspectiva de negar el dualismo deja la puerta abierta a consecuencias terribles. Si «solo somos máquinas», ¿qué sucede con el libre albedrío y la responsabilidad? ¿Cómo pueden tener sentido nuestras vidas si no somos más que una enorme colección de proteínas y otras moléculas agitándose según las leyes de la física y la química? Si los principios morales no son más que excrecencias creadas por hordas de nanomáquinas microbiológicas que medran entre nuestras orejas, ¿cómo pueden indicarnos nada que valga la pena respetar?

Crick hizo todo lo que pudo para que esta «hipótesis asombrosa» no fuera solo comprensible, sino también atractiva para el público general, pero a pesar de su prosa elegante y enérgica y de una dignidad sin par, no obtuvo grandes resultados. Creo que, en buena medida, esto fue culpa de que, a pesar de la advertencia del título, subestimó la confusión emocional que provoca esta idea. Crick era un magnífico divulgador científico, pero en este campo los problemas pedagógicos son de otra índole; no basta con captar y mantener la atención de una audiencia algo apabullada e intimidada y convencerla con una pizca de matemáticas. Cuando surge el tema de la conciencia, lo difícil es mantener a buen recaudo las angustias y las sospechas que acechan a la gente (incluidos muchos científicos) y les hacen tergiversar lo que sabemos y lanzar ataques preventivos contra las ideas peligrosas que se avecinan. Además, en este tema, todo el mundo es un experto; la gente no tiene ningún problema en que se le expliquen las propiedades químicas del calcio o los detalles microbiológicos del cáncer, pero creen que tienen una autoridad personal especial acerca de la naturaleza de sus propias experiencias conscientes,

autoridad que puede vencer a cualquier hipótesis que consideren inaceptable.

Crick no está solo. Muchos otros han probado suerte para suturar lo que uno de los mejores, Terence Deacon, ha denominado «la herida cartesiana que ha amputado la mente del cuerpo en el nacimiento de la ciencia moderna» (2011, p. 544). Sus esfuerzos han sido encomiables, fascinantes, informativos y persuasivos, pero ninguno ha conseguido ser totalmente convincente. Yo mismo he dedicado medio siglo, toda mi vida académica, a este proyecto en una docena de libros y centenares de artículos que trataban sobre diversas piezas del rompecabezas, pero no he conseguido llevar a muchos lectores de un agnosticismo receloso a una convicción relajada. Pero, impertérrito, lo intento una vez más, y ahora con la historia completa.

¿Por qué pienso que vale la pena intentarlo? En primer lugar, creo que hemos logrado unos enormes avances científicos en los últimos veinte años; actualmente, muchas de las sospechas e intuiciones de tiempos pasados pueden sustituirse por resultados bien fundamentados. Mi intención es apoyarme con firmeza en los abundantes trabajos teóricos y experimentales que se han publicado recientemente. En segundo lugar, creo que ahora entiendo mucho mejor las diversas reticencias subyacentes que encorsetan nuestra imaginación, de modo que procuraré dejarlas al descubierto y desbaratarlas a medida que avancemos, y así, por primera vez, los incrédulos podrán tomar en serio la idea de una teoría científica y materialista de sus propias mentes.

LA GRAVEDAD CARTESIANA

A lo largo de los años, recorriendo fatigosamente el campo de batalla y participando en incontables escaramuzas, he llegado a ver que hay poderosas fuerzas que actúan deformando la imaginación (incluida la mía) y que nos empujan hacia un lado y luego hacia otro. Si también aprendéis a ver estas fuerzas, descubriréis que, de golpe, todo empieza a encajar de un modo diferente. Podréis identificar las fuerzas que tiran de vuestros pensamientos y así establecer alarmas para avisaros y parachoques para protegeros, de modo que podáis resistirlas eficazmente;

al mismo tiempo, os podréis aprovechar de ellas, porque estas fuerzas no solo distorsionan, también pueden potenciar la imaginación e impulsar el pensamiento hacia nuevas alturas.

Hace más de treinta años, en una noche estrellada, estaba contemplando el cielo con algunos de mis estudiantes de la Universidad Tufts mientras mi amigo, el filósofo de la ciencia Paul Churchland, nos enseñaba a ver el plano de la eclíptica; es decir, cómo mirar a los otros planetas visibles en el cielo e imaginarlos, junto con nosotros, girando alrededor del Sol, todos sobre el mismo plano invisible; va bien ladear un poco la cabeza y recordar dónde tiene que estar el Sol, detrás de ti. De golpe, la orientación encaja y ¡tachán! ¡lo ves![4] Evidentemente, todos sabíamos desde hacía tiempo cuál es la situación de nuestro planeta en el sistema solar, pero hasta que Paul nos lo hizo ver, era poco más que una información indiferente. Inspirándome en su ejemplo, presentaré algunas experiencias reveladoras que os abrirán los ojos (o la mente, más bien) y que espero que conduzcan vuestra mente a lugares nuevos y maravillosos.

La primera fuerza deformadora, que denominaré gravedad cartesiana, es en realidad el origen de muchas otras fuerzas, a las cuales os someteré una y otra vez, en formas diferentes, hasta que también podáis verlas con claridad. Sus manifestaciones más «visibles» son familiares para la mayoría de personas; o incluso demasiado familiares, pues tendemos a creer que ya las tenemos dominadas. Las estamos subestimando; debemos mirar más allá de ellas para ver la manera en que esculpen nuestros pensamientos.

Empecemos volviendo a la «hipótesis asombrosa» de Crick. Los que insistimos en que no hay nada asombroso en ella alimentamos nuestra confianza recordando la majestuosa variedad de rompecabezas bien resueltos, descubrimientos bien perseguidos y teorías bien confirmadas de la ciencia materialista moderna y que actualmente damos por sentados. Si os paráis a pensarlo, es sorprendente todo lo que los humanos hemos descubierto en los pocos siglos desde la época de Descartes. Sabemos cómo están formados los átomos, cómo interactúan los elementos químicos, cómo se multiplican las plantas y los animales, cómo se multiplican y se propagan los patógenos microscópicos, cómo se desplazan los continentes, cómo se forman los huracanes y muchísimas cosas más. Sabemos que nuestros cerebros están hechos de los mismos ingredientes que todas las otras cosas que hemos explicado y que nosotros pertenecemos a un linaje que se puede hacer remontar al origen de la

vida. Si podemos explicar conceptos como la autorreparación de las bacterias, la respiración de los renacuajos y la digestión de los elefantes, ¿por qué el pensamiento consciente en Homo sapiens no habría de divulgar también sus secretos ante ese mismo gigante científico, que nunca deja de mejorar y reforzarse?

Se trata de una pregunta retórica, e intentar responder preguntas retóricas en lugar de acobardarse ante ellas es una buena costumbre. Así pues, ¿podría la conciencia ser un reto más desafiante que la autorreparación, la respiración o la digestión? Y si así fuera, ¿por qué? Quizá sea porque la conciencia parece tan diferente, tan privada, tan accesible íntimamente a cada uno de nosotros de una forma diferente a cualquier otro fenómeno de nuestro cuerpo. Actualmente no es muy difícil imaginar cómo funciona la respiración aun sin saber los detalles: inspiramos aire, que sabemos que es una mezcla de diferentes gases, y expiramos lo que no podemos usar (dióxido de carbono, como sabe la mayoría de la gente); de alguna forma, los pulmones deben de filtrar y recoger lo que el cuerpo necesita (oxígeno) y expulsar el resto (dióxido de carbono). La idea general no es difícil de captar. En cambio, el fenómeno de oler una galleta y, de golpe, recordar un momento concreto de nuestra infancia no parece mecanicista en absoluto. «¡Hazme una máquina de la nostalgia!» «¿Cómo? ¿Qué tendría que hacer cada una de sus partes?» Incluso los materialistas más recalcitrantes tendrán que admitir que no tienen más que unas ideas vagas y borrosas sobre cómo la actividad del cerebro puede llegar a formar la nostalgia, la melancolía o la curiosidad morbosa, por poner otro ejemplo.

Muchos pueden aceptar que no se trata tanto de una hipótesis asombrosa como de una hipótesis que nos deja perplejos y acerca de la cual lo máximo que podemos hacer es abandonar toda esperanza. Aun así, es una postura que no resulta incómodo sostener y nos puede tentar a considerar que todos aquellos que difieren (los autoproclamados Defensores de la Conciencia frente a la Ciencia) sufren algún defecto vergonzoso: narcisismo («me niego a que mi magnífica mente quede presa en manos de la ciencia»), miedo («si mi mente solo es mi cerebro, no seré responsable de nada, ¡la vida no tendrá ningún sentido!») o arrogancia («estos reduccionistas ingenuos, presos del cientificismo… no tienen ni idea de cuán lejos están de apreciar el mundo del significado con sus patéticos intentos»).

A menudo este diagnóstico es correcto. No escasean los ridículos berridos de muchos de estos Defensores, pero es cierto que las inquietudes que los

motivan no son fantasías frívolas; los que consideran que la hipótesis de Crick no es solo asombrosa, sino profundamente repulsiva, plantean una idea digna de tener en cuenta. Tampoco escasean los filósofos y científicos antidualistas que no se sienten cómodos con el materialismo y buscan algo a medio camino, algo con lo que realmente poder avanzar en la ciencia de la conciencia sin caer en ninguno de ambos extremos. El problema es que tienen una cierta tendencia a describirlo equivocadamente, exagerándolo hasta convertirlo en algo profundo y metafísico.[5]

Lo que sienten es una forma de pensar, un hábito adquirido tan bien atrincherado en nuestra psicología que negarlo o abandonarlo es impensable. Un síntoma de ello es que la confiada postura científica expresada por el «otro bando» empieza a estremecerse cuando los científicos se acercan a determinados temas sobre la conciencia y pronto se encuentran, muy a su pesar, adoptando el punto de vista de los Defensores. Para empezar, describiré este proceso metafóricamente para ofrecer un marco a partir del cual comprender lo que está pasando, de un modo menos metafórico, más explícito y factual.

Supongamos que el aspirante a esclarecedor de la mente empieza con su propia mente. Ahí está, en su casa, en el planeta Descartes, meditando sobre la tarea a la que se enfrenta y observando el universo desde su «punto de vista en primera persona». Desde este mirador, se sirve de todos los elementos familiares de su mente para mantener el rumbo, mientras que la gravedad cartesiana es la fuerza que lo mantiene en este punto de vista egocéntrico «desde dentro». Recordando a Descartes, su soliloquio podría ser algo como: «Heme aquí, una cosa pensante consciente, íntimamente familiarizada con las ideas de mi mente, que conozco mejor que nadie, simplemente porque son las mías». No puede evitar ser un Defensor de su propio hogar. Mientras tanto, desde algún lugar lejano, llega el explorador científico de la conciencia, que se acerca confiadamente al planeta Descartes, armado con instrumentos, mapas, modelos y teorías, y prepara su pronta conquista. Sin embargo, cuanto más se acerca, más incómodo se siente; es arrastrado hacia una orientación que sabe que tiene que evitar, pero la fuerza es demasiado intensa. En cuanto aterriza en el planeta Descartes se da cuenta de que, súbitamente, su orientación se ha invertido, pasando a una orientación en primera persona, con los pies en el suelo, pero incapaz de alcanzar o usar los instrumentos que había traído para completar su conquista. La gravedad cartesiana es irresistible cuando uno se acerca tanto a la

superficie del planeta Descartes. ¿Cómo llegó ahí y qué sucedió en esa confusa inversión en el último momento? (las inversiones extrañas serán uno de los temas principales de este libro). Parece haber dos orientaciones en competición: el punto de vista en primera persona de los Defensores y el punto de vista en tercera persona de los científicos, no muy diferentes de las dos formas de ver las ilusiones ópticas favoritas de los filósofos, el conejo/pato y el cubo de Necker. Es imposible adoptar ambos puntos de vista a la vez.

A veces, el problema que plantea la gravedad cartesiana se denomina vacío explicativo o laguna explicativa (Levine 1983), pero las discusiones en torno a esta idea me parecen básicamente infructuosas, porque los participantes acostumbran a verla como un abismo, no como un fracaso de sus imaginaciones. Puede que hayan descubierto el vacío, pero no son capaces de ver lo que realmente es porque no se han preguntado «cómo llegó a ser de esta manera». Al reconsiderar el vacío como un elemento dinámico distorsionador de la imaginación, que ha surgido por un buen motivo, podemos aprender cómo atravesarlo o, lo que sería lo mismo, hacerlo desaparecer.

La gravedad cartesiana, a diferencia de la gravedad de la física, no actúa sobre los objetos proporcionalmente a sus masas y a la proximidad a otros objetos con masa; actúa sobre las ideas o representaciones de cosas en proporción a su proximidad, en términos de contenido, a otras ideas que tienen un papel privilegiado en el mantenimiento de un ser vivo (espero que el significado de esto vaya quedando claro poco a poco, para así abandonar esta forma metafórica de hablar, al igual que una escalera que nos ha servido para encaramarnos a lo alto pero que ya no necesitamos). La idea de la gravedad cartesiana, tal como la he presentado, solo es una metáfora, pero el fenómeno al cual da nombre es perfectamente real, una fuerza perturbadora que acosa a nuestra imaginación (aunque a veces la ayuda) pero que, a diferencia de la gravedad de la física, es un fenómeno producto de la evolución. Para entenderla, tendremos que preguntarnos cómo surgió en la Tierra.

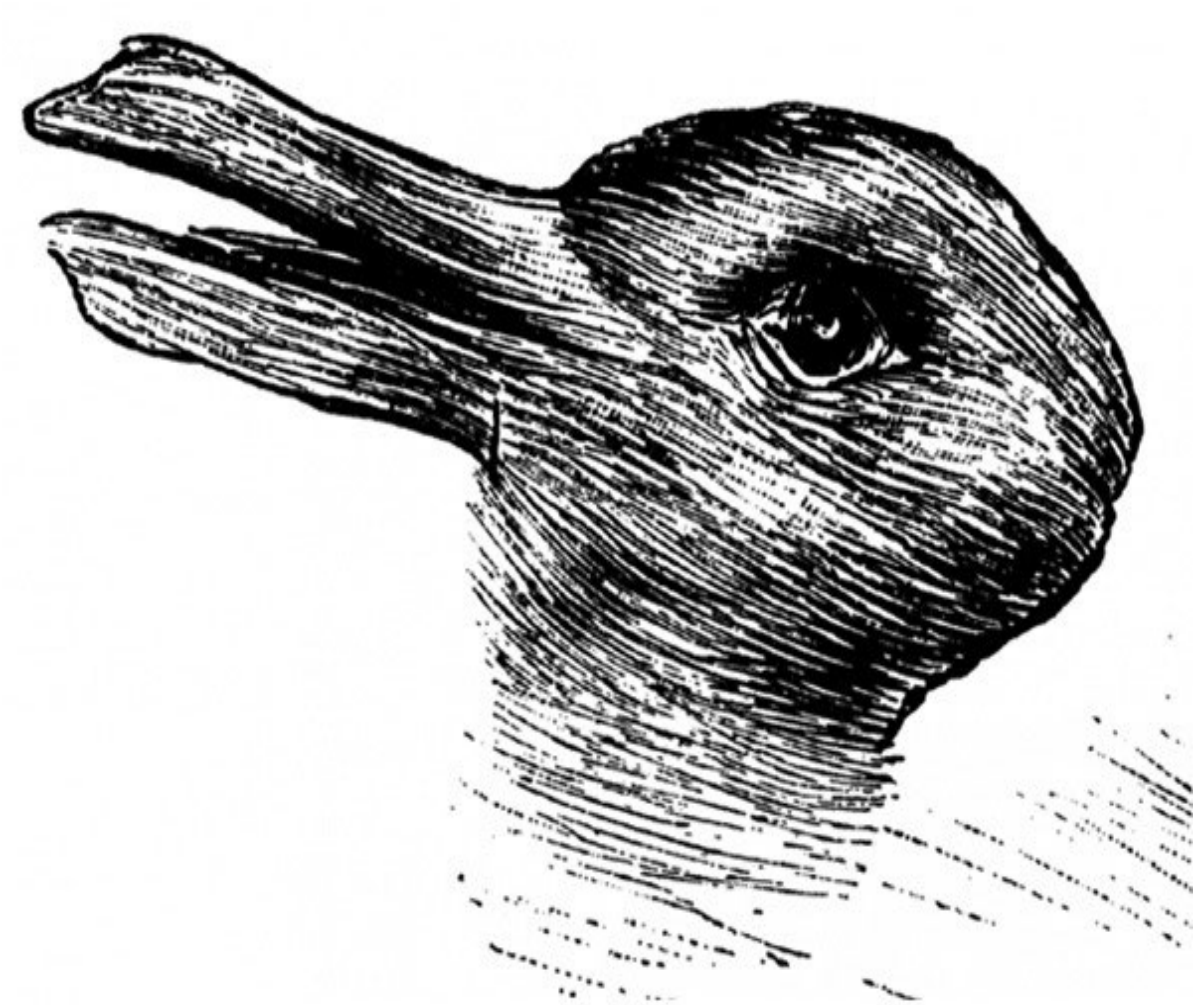


FIGURA 1.1. El conejo/pato

Para responder a esta pregunta tendremos que pasar varias veces por la misma historia, subrayando cada vez aspectos diferentes. Solemos subestimar la intensidad de las fuerzas que distorsionan nuestra imaginación, especialmente al enfrentarnos a percepciones irreconciliables que son «innegables». No es que no podamos negarlas, es que no vamos a hacerlo, ni tan solo lo intentaremos. Practicando con las fuerzas que nos resultan fáciles de identificar (el especifismo, el excepcionalismo humano, el machismo) nos prepararemos para identificar fuerzas más sutiles. En el siguiente capítulo trataré brevemente el origen de la vida en el planeta y esbozaré a grandes rasgos la historia que vendrá a continuación; también abordaré una de las primeras objeciones que supongo que se le ocurrirán al lector mientras lea ese

esbozo. Planteo los procesos evolutivos como procesos de diseño (procesos de investigación y desarrollo, de I+D), pero desde hace tiempo, esta perspectiva adaptacionista, o de ingeniería inversa, se ha visto envuelta por unas sospechas inmerecidas; tal como veremos, al contrario de la creencia generalizada, el adaptacionismo continúa sano y salvo dentro de la biología evolutiva.

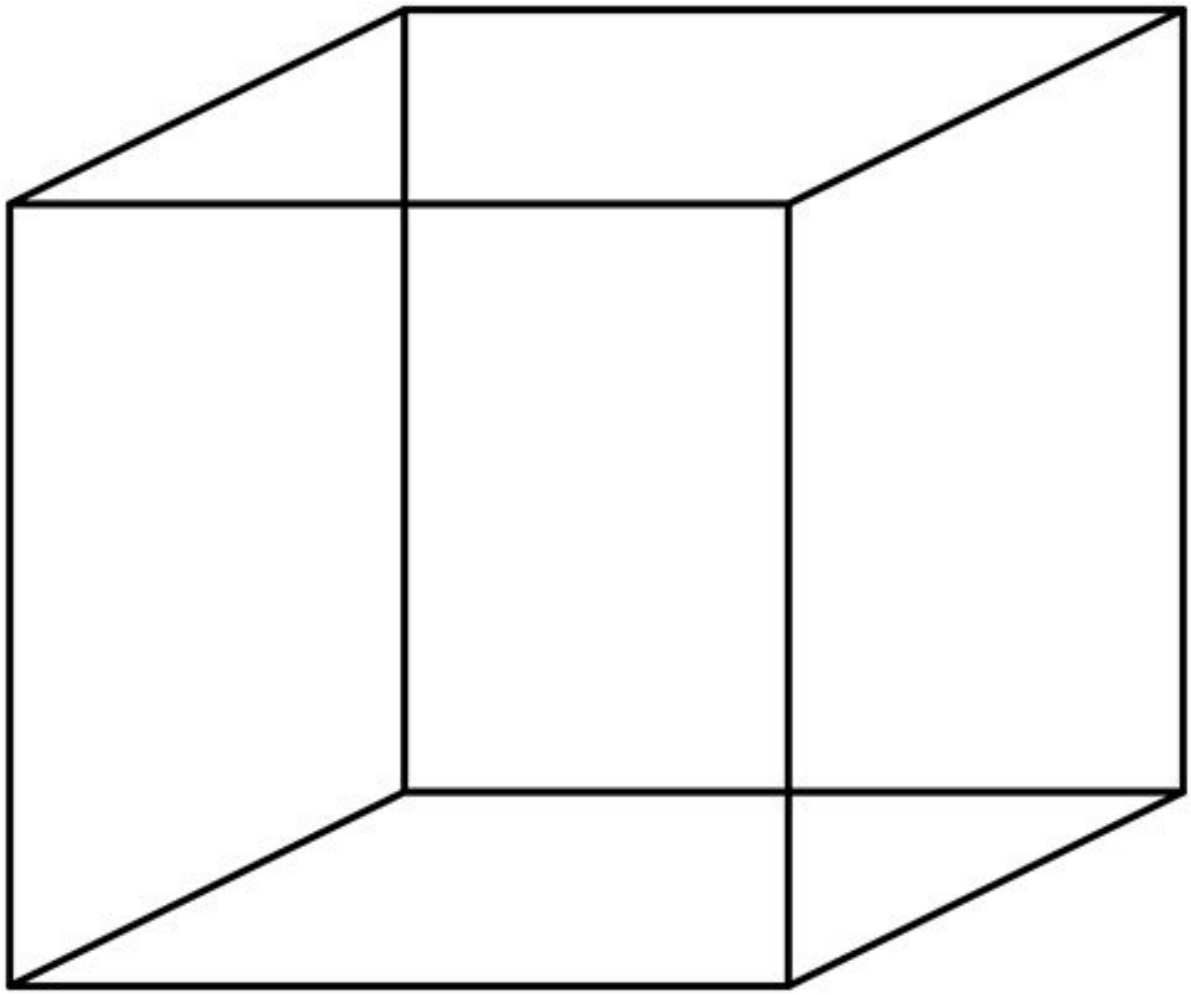


FIGURA 1.2. El cubo de Necker

2

ANTES DE LAS BACTERIAS

Y ANTES DE BACH

¿POR QUÉ BACH?

Para lograr una buena perspectiva de nuestra historia, tendremos que retroceder en el tiempo hasta una época anterior a las bacterias, antes de que surgiera cualquier tipo de vida, porque algunos de los requisitos para la existencia misma de la vida reverberan a través de los eones y permiten explicar rasgos de las mentes actuales. Pero antes de abordar esta historia, dadme un momento para remarcar una palabra: «Bach». Podría haber escogido como título «De las arqueas a Shakespeare» o «De E. coli a Einstein» o también «De los procariotas a Picasso», pero lo cierto es que la aliteración en «De las bacterias a Bach» me pareció irresistible.

Por cierto, ¿qué hay del hecho de que todos los candidatos que acabo de considerar en mi panteón de grandes mentes son hombres? ¡Vaya tropezón más embarazoso, justo al empezar nuestro viaje! Realmente, ¿quiero alienar a tantos lectores desde un buen principio? ¿En qué estaba pensando? Lo cierto es que lo he hecho a propósito, para ofrecer un ejemplo simple y comparativamente moderado de uno de los tipos de fuerza de gravedad cartesiana con la que nos enfrentaremos. Si os habéis sentido molestos al daros cuenta de que mi lista de genios es exclusivamente masculina, vamos por buen camino; significa que no vais a olvidar que os he robado un poco de vuestra paciencia, que os tendré que devolver más adelante en este libro. El enojo (como cualquier otra reacción emocional intensa, desde el miedo a la diversión) es una especie de recordatorio, que subraya en vuestra memoria ese elemento insultante

e impide que lo olvidéis. En este punto, os tengo que pedir que resistáis el impulso de lanzar un ataque preventivo; durante nuestro viaje tendremos que identificar hechos incómodos sin poder darnos el gusto de avanzar explicaciones o refutaciones prematuras. Aunque me encanta que mis lectores no solo me presten atención, sino que, además, vayan por delante de mis propios pensamientos, ahora preferiría que esperaseis vuestro momento y que no fuerais tan duros (o que me dieseis un poco más de cuerda para ahorcarme, si preferís verlo así), en lugar de intentar desbaratar rápidamente con vuestros presentimientos mis intentos de ofrecer una descripción tranquila y objetiva.

Así pues, hagamos una pausa y revisemos unos pocos hechos, dejando las explicaciones y las refutaciones para más adelante. Es un hecho evidente que, aunque ha habido muchas mujeres brillantes, responsables de grandes logros, ninguna ha logrado la misma categoría icónica que Aristóteles, Bach, Copérnico, Dickens o Einstein, entre otros. Podría añadir fácilmente una docena más de hombres de la misma altura; ahora intentad añadir a alguna gran pensadora que pudiera sustituir a cualquiera de estos hombres para el papel emblemático de mi título (mis favoritas serían Jane Austen, Marie Curie, Ada Lovelace e Hipatia de Alejandría; creo que no he omitido a ninguna de las candidatas más claras, pero el tiempo lo dirá).

No ha habido ningún supergenio femenino, todavía. ¿Qué es lo que podría explicar esta situación? ¿Opresión política? ¿Las profecías machistas autocumplidas que privan a las chicas de modelos inspiradores? ¿Los prejuicios del entorno, a lo largo de los siglos? ¿Los genes? No adelantéis ninguna conclusión, aunque creáis que la respuesta es obvia (yo no lo creo). Pronto veremos que los genes, aun siendo un personaje principal en la historia de la mente, no son ni mucho menos tan importantes como muchos creen; los genes pueden explicar las capacidades animales básicas, pero no pueden explicar la genialidad. Además, la idea tradicional de que la grandeza de las sociedades humanas se debe a la brillantez creativa de (algunos de) sus habitantes es, a grandes rasgos, ingenua, como intentaré mostrar. La cultura humana, por ella misma, es un generador de brillantes innovaciones más fecundo que cualquier grupo de genios, sean del sexo que sean. Y esto ocurre gracias a un proceso de evolución cultural que es tan «autor» de nuestros logros más espectaculares como lo pueda ser cualquier gran pensador.

La mera idea de que la evolución por selección natural pueda tener un papel básico en la comprensión de la cultura humana resulta odiosa para muchas personas, incluso personas razonables y sensatas. Consideran que la cultura humana es algo trascendente, el don milagroso que distingue a los seres humanos de las bestias, el último baluarte frente a un reduccionismo insidioso, frente a un determinismo genético y frente al espíritu filisteo que ven en la ciencia actual. Del mismo modo, frente a estos «sabelotodo», para algunos científicos tenaces cualquier apelación a «cultura» apesta a una promoción del misterio, o aún peor.

Cuando oigo la palabra cultura, desenfundo mi pistola.[1]

Debo pedir ahora a «ambos bandos» que enfunden sus pistolas y tengan un poco de paciencia. Hay un punto medio que puede hacer justicia simultáneamente a las humanidades y a la ciencia, explicando cómo surge y avanza la cultura humana mediante un proceso de evolución de elementos culturales (los memes) que invadieron los cerebros humanos del mismo modo que los virus invaden los cuerpos humanos. En efecto, no se ha demostrado que el concepto de meme sea una mala idea y en este libro también subirá al estrado. Aquellos en ambos bandos que se burlan y se ríen de los memes podrán ver que, ciertamente, hay buenas objeciones a este concepto (además de las «refutaciones» defectuosas que muchos que no soportan la idea han aceptado acríticamente), pero estas objeciones nos llevan a mejoras que, al fin y al cabo, rescatan el concepto de meme.

Así pues, ¿en qué bando estoy? Los lectores que quieran enmarcar la discusión en estos términos no han entendido la cuestión. Esta polarización de los puntos de vista, con espectadores vitoreando y abucheando a los combatientes, no es más que una primera manifestación, especialmente obvia, de las fuerzas que intento visibilizar y neutralizar. Nos quedan muchas más por delante: presiones más sutiles e insidiosas que afectan al pensamiento de científicos, filósofos y profanos.

Volvamos ahora a mi primer recorrido por la historia.

EL MUNDO PREBIÓTICO COMO UNA PARTIDA DE AJEDREZ

La primera y más simple forma de vida capaz de reproducirse, no muy diferente de una bacteria, ya era un sistema autosuficiente brillantemente diseñado y sorprendentemente complejo (un momento… ¿no acabo de ayudar y reconfortar a los seguidores del Diseño Inteligente? Mi respuesta es que no. Pero, a ver, ¿cómo puede ser que un buen darwinista, materialista y ateo como yo, ni se inmute al decir que las primeras formas de vida capaces de reproducirse estaban brillantemente diseñadas? Aguantad un momento, no disparéis todavía).

Un conocido problema del tipo del «huevo y la gallina», apreciado por la gente del Diseño Inteligente, nos reta con la siguiente «paradoja» del origen de la vida: la evolución por selección natural no podía empezar hasta que hubiera elementos que se reprodujeran, ya que en caso contrario no habría descendencia para heredar los mejores diseños; pero resulta que la cosa más simple que se puede reproducir es demasiado compleja para haber aparecido por azar.[2] En consecuencia, afirman, la evolución no se podría haber iniciado sin la ayuda de un Diseñador Inteligente. Se trata, como veremos, de un razonamiento defectuoso, una combinación de error y fracaso de la imaginación. Ahora bien, sí que debemos reconocer el hecho sorprendente de que la primera mezcla de moléculas capaz de reproducirse a sí misma con fiabilidad era (tiene que haber sido) una maravilla de ingeniería, formada por miles de partes complejas trabajando conjuntamente.

Los investigadores que trabajan en el origen de la vida se enfrentan a un reto muy evidente: ¿cómo puede haber sucedido esto sin un milagro? (quizá el responsable fuera un diseñador inteligente de otra galaxia, pero esto solo pospondría la pregunta y la haría más difícil de resolver). La forma de proceder es clara: empezar con las especificaciones mínimas para una «cosa» viva y capaz de reproducirse (es decir, una lista de todo lo que tiene que poder hacer) e ir hacia atrás, haciendo un inventario de las materias primas disponibles (denominadas, a menudo, moléculas básicas de la química prebiótica) y preguntándose qué posible secuencia de sucesos podría reunir, sin milagros, todas las piezas necesarias en las posiciones adecuadas para lograr la tarea. Tened en cuenta que estas

especificaciones mínimas indican lo que tal cosa debe hacer (una lista de funciones), no sus partes ni materiales. Una ratonera tiene que atrapar ratones y un abrelatas tiene que abrir latas; una cosa viva tiene que capturar energía y protegerse (o repararse) durante el tiempo necesario para reproducirse.

¿Cómo podría surgir esta cosa viva? Si podéis responder esta pregunta, ¡premio! Sería algo así como un jaque mate en una partida de ajedrez. Es una tarea formidable, con muchas lagunas a cubrir, pero cada año se producen avances alentadores y la sensación es que la tarea puede completarse, que la partida puede ganarse. Hay muchas maneras en que puede haber surgido la vida a partir de elementos no vivos, pero solo con encontrar una que sea digna de confianza científica (hasta descubrir una alternativa mejor) sofocaría las voces de «es imposible en principio» de una vez por todas. Sin embargo, encontrar solo una de estas maneras es una tarea tan abrumadora que ha alimentado la convicción entre muchos investigadores de que, aunque los procesos que deben invocarse para crear el producto final son completamente ciegos y sin propósito alguno, tal producto no es solamente complejo, sino increíblemente eficaz en lo que hace; un diseño brillante, en suma. Necesitamos todo el ingenio humano posible para aplicar ingeniería inversa y comprender cómo se ensambló la «cosa». Un comentario de Jack Szostak, uno de los principales investigadores de uno de los avances más relevantes de los últimos años (Powner, Gerland y Sutherland 2009), ilustra perfectamente esta postura (no os preocupéis por los detalles químicos que aparecen; basta ver cómo se realiza la investigación, tal como indico con los fragmentos en cursiva):

Durante cuarenta años, los esfuerzos para comprender la síntesis prebiótica de los ladrillos ribonucleotídicos de ARN se han basado en la suposición de que tienen que haberse formado a partir de sus tres componentes moleculares: una base nucleotídica (que puede ser adenina, guanina, citosina o uracilo), una ribosa (un azúcar) y un grupo fosfato. Entre las muchas dificultades que encontramos los que nos dedicamos a este tema, la más frustrante ha sido la incapacidad de hallar cualquier forma de unir adecuadamente las bases pirimidínicas (citosina y uracilo) con la ribosa […] Pero Powner y sus colaboradores resucitan las expectativas del modelo de «ARN primero» explorando una ruta para la

síntesis de los ribonucleótidos pirimidínicos en que el azúcar y la base surgen a partir de un precursor común. En esta ruta, la estructura ribonucleotídica completa se forma sin necesidad de usar moléculas libres de azúcar y de las bases como intermediarios. Esta intuición básica, combinada con un conjunto de innovaciones adicionales, proporciona una solución notablemente eficaz al problema de la síntesis prebiótica de los ribonucleótidos.

Greg Mayer (2009), un biólogo evolutivo, aporta un comentario relevante sobre este tema:

John Sutherland, uno de los coautores del trabajo de Powner, y en cuyo laboratorio se llevó a cabo la investigación, trabajó en el problema durante doce años antes de hallar la solución. ¿Qué hubiera pasado de haber abandonado después de diez años? ¿Podríamos haber llegado a la conclusión de que tal síntesis era imposible? No. Este trabajo demuestra la futilidad de las diversas argumentaciones que se basan en la ignorancia como premisa principal (el argumento del diseño o teleológico, el argumento del dios tapagujeros, el argumento de la incredulidad personal).

En este libro me basaré en el punto de vista de la ingeniería inversa, suponiendo que cada ser vivo es el producto de procesos físicos sin misterio alguno que, gradualmente, unieron todos los elementos necesarios, perfeccionándolos a lo largo del camino y, finalmente, produjeron el sistema que observamos o bien un sistema intermedio hipotético, un peldaño que sería un progreso claro hacia los seres vivos que existen. La cascada de procesos debe efectuar cambios que, retrospectivamente, podamos considerar que son mejoras en el diseño de los sistemas que se están formando (nos acercamos a un jaque mate… ¿estamos avanzando?). Hasta la aparición de sistemas que pudieran denominarse estrictamente sistemas reproductores, los procesos en marcha eran solo protoevolutivos, semidarwinianos, análogos parciales de la evolución por selección natural; eran procesos que aumentaban la probabilidad de que surgieran y persistieran diversas combinaciones de

ingredientes, que concentraban la materia prima molecular, hasta que este proceso llevó, finalmente, al origen de la vida. Un ser vivo debe capturar la cantidad necesaria de energía y de materiales y esquivar su propia destrucción durante el tiempo suficiente como para poder construir una réplica de sí mismo lo suficientemente buena. El punto de vista de la ingeniería inversa es ubicuo en biología y es obligatorio en las investigaciones sobre el origen de la vida, y siempre involucra algún tipo de consideración de optimización: ¿cuál es la estructura química más simple que podría hacer x? o ¿es el fenómeno x lo bastante estable para sustentar el proceso y?

En un ensayo muy influyente, Stephen Jay Gould y Richard Lewontin (1979) acuñaron el término paradigma panglossiano, deliberadamente ofensivo, para referirse al adaptacionismo, la rama de la biología que se basa en el principio metodológico de suponer, hasta que se demuestre lo contrario, que todas las partes de un organismo sirven para algo; es decir, que tienen un papel útil que representar, como bombear sangre, mejorar la locomoción, eludir infecciones, digerir alimentos, disipar calor, atraer a la pareja, etc. Esta suposición forma parte del punto de vista de la ingeniería inversa, que considera que todos los seres vivos están formados, de modo eficaz, por partes con funciones (hay excepciones bien aceptadas, como rasgos que una vez fueron útiles para algo, pero que actualmente son vestigiales y que se mantienen a no ser que sea demasiado caro mantenerlos, así como rasgos que no tienen ninguna función real y solo «derivaron» hasta «quedar fijados» por azar).

La broma de Gould y Lewontin es, en realidad, una caricatura reciclada. En Cándido, Voltaire creó al doctor Pangloss, una maliciosa y divertida caricatura del filósofo Leibniz, que sostenía que nuestro mundo es el mejor de todos los mundos posibles. Según la desbocada imaginación del doctor Pangloss, no existían rarezas, deformidades ni catástrofes de la naturaleza que no se pudieran considerar, retrospectivamente, una bendición que un dios benevolente dispondría para nosotros, afortunados habitantes de un mundo perfecto. La sífilis, por ejemplo, «es indispensable en este mundo, el mejor de todos. Porque si Colón, al visitar las Indias Occidentales, no hubiera contraído esta enfermedad, que envenena la fuente de la generación, que a menudo la impide y se opone claramente al gran fin de la Naturaleza, no tendríamos chocolate ni cochinilla» (citado por Gould y Lewontin, 1979). Los estudiosos de Leibniz, con algo de razón, afirman que la parodia de Voltaire es muy

injusta para el filósofo, pero dejemos este tema. ¿Fue la reutilización de la idea por parte de Gould y Lewontin una caricatura injusta del uso de las suposiciones de optimización en biología? Sí, y además ha tenido dos efectos desafortunados: en primer lugar, su ataque al adaptacionismo ha sido mal interpretado por algunas personas temerosas de la evolución como una refutación de la teoría de la selección natural; en segundo lugar, ha convencido a muchos biólogos de que no solo deben censurar su lenguaje, sino también su pensamiento, como si la ingeniería inversa fuera una especie de truco ilícito que deberían evitar a toda costa.

Los que trabajan en el origen de la vida han ignorado esta crítica «panglossiana» de sus métodos, a sabiendas de que sus suposiciones estratégicas sirven para guiar la investigación y alejarla de divagaciones infructuosas. No tiene sentido considerar reacciones químicas que no pueden generar de ninguna manera una estructura que se supone que es un componente necesario. Sin duda, hay cierto riesgo en esta estrategia; tal como señala Szostak, durante años los investigadores han cometido el error de suponer que la manera más eficaz de unir las bases nitrogenadas con la ribosa era directamente, y pasaron por alto la vía más tortuosa en que el ribonucleótido surge a partir de un precursor común, sin los pasos intermedios que habían parecido necesarios.

En el ajedrez, un gambito es una estrategia en la que se sacrifica material para conseguir una ventaja (como retirarse para avanzar desde una mejor posición). Al intentar calcular lo que va a hacer tu oponente, los gambitos son difíciles de detectar, porque a primera vista parecen malos movimientos que se pueden ignorar sin problema, ya que tu oponente no es tan tonto. El mismo riesgo de ignorar sendas tortuosas pero fructíferas acecha a los que aplican la ingeniería inversa en biología, pues, como dijo Francis Crick, enunciando lo que llamó la Segunda Regla de Orgel, «la evolución es más lista que tú». La asombrosa forma en que la fábrica ciega y sin propósito de la evolución (incluyendo la evolución química prebiótica) descubre soluciones disparatadas a los problemas no es ninguna prueba de un Diseñador Inteligente, ni tampoco es ninguna justificación para abandonar la ingeniería inversa, lo que implicaría abandonar totalmente nuestra búsqueda; al contrario, es una justificación para persistir y mejorar nuestra partida de ingeniería inversa. Como en el ajedrez, no abandones, aprende de tus errores y sigue explorando, tan imaginativamente como puedas, y siempre teniendo presente que tus

hipótesis, por muy plausibles que parezcan, pueden ser refutadas, situación que debe buscarse a conciencia.

He aquí un ejemplo de un posible gambito en el origen de la vida. Al principio, es tentador suponer que el primer ser vivo capaz de reproducirse tiene que haber sido el ser vivo más simple posible (de acuerdo con las condiciones existentes en el planeta en ese momento). Empecemos por el principio: construye el replicador más simple que puedas imaginar y, luego, continúa a partir de estos cimientos. Pero esto no es en absoluto necesario. Es posible, y yo creo que más probable, que el primer replicador real fuera un aglomerado de objetos diversos, poco elegante, complicado, lento y costoso, una especie de invento del tebeo digno del Profesor Franz de Copenhague; y una vez la bola de nieve de la replicación empezó a rodar, este torpe replicador se fue simplificando una y otra vez, en competición con los demás. Muchos de los más enigmáticos trucos de magia se basan en que los espectadores ni se imaginan los esfuerzos absurdamente extravagantes que llegan a hacer los magos para conseguir un efecto desconcertante. Si deseáis aplicar ingeniería inversa a los trucos de magia, deberíais tener muy presente que los magos no tienen reparos en hacer enormes y estrafalarios esfuerzos para conseguir pequeños efectos, de los que se pueden aprovechar. Del mismo modo, la naturaleza tampoco tiene reparos… ni restricciones de dinero, y, además, tiene todo el tiempo del mundo.

Hablar de progreso o de mejoras en el lento e incierto proceso de la biogénesis no es dejarse llevar por juicios de valor ilícitos (que, efectivamente, no tienen sitio en la ciencia); es más bien reconocer los siempre presentes requisitos de estabilidad y eficacia para cualquier ser vivo. Si queréis, podéis imaginar a un grupo de bioquímicos investigando cómo puede aparecer algo increíblemente terrible, una máquina del juicio final o un rayo de la muerte autorreplicante. Aun así, tendrían que acotar su búsqueda imaginando posibles rutas para construir ese horror; y podrían maravillarse ante la genialidad de su diseño, una vez descubierto. Más adelante comentaré más cosas acerca de las suposiciones e implicaciones de la ingeniería inversa en biología, pero aquí espero frustrar una refutación prematura de mi proyecto por parte de aquellos que hayan sido persuadidos, directamente o de oídas, de que la propaganda de Gould y Lewontin contra el adaptacionismo era un golpe mortal. A diferencia de la opinión ampliamente extendida por su famosa obra, el adaptacionismo continúa sano y salvo; la ingeniería inversa,

cuando se aplica con la debida atención a los riesgos y a las obligaciones que conlleva, aún es la ruta principal para llegar a descubrimientos en biología, y la única en el desafiante mundo de la química prebiótica relacionada con el origen de la vida.[3]

A continuación, me gustaría echar un vistazo al fenómeno del origen de la vida desde una perspectiva más filosófica, de acuerdo con el origen de las razones. ¿Existe diseño en la naturaleza o es solamente una apariencia de diseño? Si consideramos que la biología evolutiva viene a ser una ingeniería inversa, ¿implica ello que hay razones para la disposición de las partes de los seres vivos? ¿Cuáles? Razones o motivaciones, ¿de quién? ¿O quizá puede haber razones sin razonador? ¿Diseño sin diseñador?

3

SOBRE EL ORIGEN DE LAS RAZONES

¿MUERTE O RESURRECCIÓN DE LA TELEOLOGÍA?

A menudo se atribuye a Darwin el derrocamiento de la doctrina aristotélica según la cual todo lo que existe en el mundo tiene un propósito, una finalidad, un fin (en el sentido en que se usa en la frase «el fin justifica los medios») o, como se dice en francés, una raison d'être (una 'razón de ser').

Aristóteles identificó cuatro preguntas que podríamos hacer sobre cualquier cosa:

De qué está hecha; es decir, su causa material.

Cuál es su estructura; es decir, su causa formal.

Cómo se originó; es decir, su causa eficiente.

Cuál es su propósito; es decir, su causa final o télica.

En la cuarta causa, el término télico procede del griego telos, del cual también deriva el término teleología. Con frecuencia se dice que la ciencia ha proscrito el telos, gracias a Darwin. Tal como bien expresó Karl Marx (1861) en su valoración de El origen de las especies, «aquí no solamente se asesta por primera vez un golpe mortal a la "teleología" en

las ciencias naturales, sino que también se explica empíricamente su significado racional».

Un análisis más detenido, sin embargo, nos revela que Marx está siendo ambiguo entre dos aspectos que se siguen planteando:

Debemos desterrar cualquier formulación teleológica de las ciencias naturales

o

ahora que podemos «explicar empíricamente» el «significado racional» de los fenómenos naturales sin recurrir a antiguas creencias (entelequias, Creadores Inteligentes y cosas por el estilo), podemos sustituir la anticuada teleología con una nueva teleología posdarwiniana.

En la actualidad, esta ambivalencia está firmemente anclada en la práctica y las intenciones de muchos científicos sensatos. Por un lado, tenemos a biólogos que, de modo habitual hacen referencia a las funciones de comportamientos como forrajear o marcar territorio, de órganos como los ojos o las vejigas natatorias, de máquinas subcelulares como los ribosomas, de ciclos químicos como el ciclo de Krebs y de macromoléculas como las proteínas motrices o la hemoglobina. Pero algunos biólogos y filósofos de la biología se muestran incómodos con estas afirmaciones e insisten en que todas estas referencias a funciones y propósitos son solo una forma de hablar, una metáfora útil, y que, estrictamente, no hay ninguna función, ningún propósito, ningún tipo de teleología en el mundo. Aquí podemos ver el efecto de otra de las fuerzas de la gravedad cartesiana que distorsionan la imaginación. El pensamiento cartesiano es tan atractivo que, para resistirse a él, algunos consideran que debemos aferrarnos al principio abstemio según el cual, siempre que haya el más mínimo riesgo de infección con conceptos precientíficos (de almas y espíritus, de teleología aristotélica y de cosas parecidas) es mejor ser precavido y establecer una cuarentena absoluta y

clarísima. En muchos casos, es un principio excelente: el cirujano que reseca un tumor aplica un generoso margen alrededor del tejido sospechoso; los políticos crean cordones de seguridad para mantener a distancia armas (o ideologías) peligrosas.

Un poco de proselitismo ayuda a que la gente esté atenta. «Paranoicos darwinistas» (Francis 2004; Godfrey-Smith 2009) o «conspiranoicos» (Rosenberg 2011) son algunos de los calificativos que se arrojan a los teleólogos impenitentes. Por supuesto, se puede defender una postura intermedia que prohíba algunos excesos teleológicos pero que admita una discusión más formal y limitada acerca de funciones; y, ciertamente, los filósofos han concebido varias de estas posturas. Mi apreciación informal es que muchos científicos asumen que existe esta saludable posición intermedia y que debe de estar respaldada por algún libro o algún artículo que leyeron años atrás; y, sin embargo, por lo que sé, no hay tal texto clásico de consenso,[1] y muchos científicos que, inocentemente, hacen referencia a las funciones de lo que sea que están estudiando señalan que nunca cometerían el pecado de la teleología.

Otra de las fuerzas que están actuando aquí es el deseo de no animar ni dar alas a los creacionistas y a toda la tropa del Diseño Inteligente. Cuando hablamos de diseño y de propósito en la naturaleza, parece que les estemos dando algo de razón; en consecuencia, algunos piensan que es mejor mantener una rígida prohibición sobre estos temas y subrayar que, en sentido estricto, no hay nada en la biosfera que haya sido diseñado (excepto lo que han diseñado los artífices humanos). La forma que tiene la naturaleza de generar sistemas complejos (órganos, comportamientos, etc.) es tan diferente de la forma de trabajar de un artesano que no deberíamos emplear el mismo lenguaje para describirlos. Es por ello por lo que Richard Dawkins habla de rasgos designoides en los organismos (a veces; p. ej. 1996) y en El cuento del antepasado (2004) afirma que «la ilusión de diseño que evoca la selección natural de Darwin es asombrosamente potente». Yo no estoy de acuerdo con esta exagerada austeridad, que puede resultar muy contraproducente. Hace pocos años oí a unos estudiantes de la facultad de medicina de Harvard charlando en un bar y se maravillaban de las complejidades de la maquinaria proteica del interior de la célula. Uno de ellos exclamó «¿Cómo puede ser que alguien crea en la evolución ante todo este diseño?» Sus compañeros no hicieron ninguna objeción, fueran cuales fueran sus ideas. ¿Cómo puede alguien decir algo semejante? Los evolucionistas no se sienten cohibidos por las complejidades de la naturaleza; al contrario, ¡se deleitan con ellas!

Descubrir y explicar la evolución de las complejidades intracelulares que gobiernan la vida de una célula ha sido uno de los mayores triunfos de la microbiología evolutiva en los últimos años. Y, sin embargo, la observación de ese estudiante sugiere que una de las ideas que están ganando terreno en la interpretación habitual es que los biólogos evolucionistas son renuentes a «admitir» o «aceptar» el evidente diseño que hay en la naturaleza. La gente debería ser más sensata, ¡especialmente los estudiantes de medicina!

Abundando sobre el tema, consideremos las declaraciones de Christoph Schönbron, arzobispo de Viena que, cautivado por el Diseño Inteligente, niega la teoría de la selección natural sobre la base de que no puede explicar todo el diseño que observamos. En un artículo de opinión en The New York Times, titulado «Finding Design in Nature» (7 de julio de 2005) afirmó tristemente que:

La Iglesia católica, si bien deja a la ciencia muchos detalles acerca de la historia de la vida en la Tierra, proclama que, a la luz de la razón, el intelecto humano puede identificar claramente propósito y diseño en el mundo natural, incluyendo el mundo de los seres vivos. Puede que la evolución sea cierta en el sentido de un ancestro común, pero no en el sentido neodarwinista (un proceso sin planificación y sin guía de variaciones al azar y selección natural). Cualquier sistema de pensamiento que niegue o intente justificar la apabullante evidencia de diseño en la naturaleza es ideología, no ciencia.

¿Qué batalla queremos librar? ¿Queremos intentar convencer al público general de que ese diseño que están viendo y que es asombrosamente obvio a cualquier escala de la biología en realidad no existe? ¿o quizá es mejor intentar convencerles de que lo que Darwin ha demostrado es que puede haber diseño (diseño real, tan real como pueda ser) sin necesidad de ningún Diseñador Inteligente? Hemos convencido a todo el mundo de que los átomos no son atómicos y de que la Tierra gira alrededor del Sol; ¿por qué deberíamos acobardarnos ante la tarea pedagógica de mostrar que puede haber diseño sin diseñador? Así pues, aquí estoy otra vez (haciendo un nuevo hincapié) para defender la siguiente postura:

La biosfera está llena a rebosar de diseño, de propósito, de razones. Lo que yo denomino enfoque de diseño o actitud de diseño predice y explica las características del mundo vivo utilizando las mismas suposiciones que funcionan tan bien cuando se aplica la ingeniería inversa a los artefactos creados por diseñadores humanos más o menos inteligentes.

Para intentar comprender, explicar y predecir los fenómenos, hay tres estrategias, actitudes o enfoques diferentes, pero estrechamente relacionados: el enfoque físico, el enfoque de diseño y el enfoque intencional (Dennett 1971, 1981, 1983, 1987 y otros). El enfoque físico es el menos arriesgado, pero también el más difícil: tratas el fenómeno en cuestión como un sistema físico, que obedece las leyes de la física, y empleas tus conocimientos en este campo para predecir lo que sucederá a continuación. El enfoque de diseño solo funciona para cosas que han sido diseñadas, ya sean artefactos o seres vivos o sus partes, y que tienen funciones o propósitos. El enfoque intencional trata, básicamente, cosas que han sido diseñadas para usar información y así cumplir sus funciones; considera la cosa como un agente racional, atribuyéndole «creencias», «deseos» y «racionalidad» y suponiendo que actuará racionalmente.

La evolución por selección natural no es, por sí misma, algo diseñado. No es un agente con propósito, pero actúa como si lo fuera (ocupa la plaza dejada vacante por el Diseñador Inteligente); es un conjunto de procesos que «hallan» y «rastrean» razones para que las cosas se dispongan de una determinada manera y no de otra. La principal diferencia entre las razones que encuentra la evolución y las razones que hallan los diseñadores humanos es que las segundas suelen estar representadas en las mentes de los diseñadores, mientras que las razones descubiertas por la selección natural solo son representadas por primera vez por los investigadores humanos que consiguen aplicar ingeniería inversa a las creaciones de la naturaleza. El título de la obra de Dawkins El relojero ciego (1986) evoca elegantemente la aparente paradoja de estos procesos: por un lado, son procesos ciegos, sin sentido, sin objetivos; pero, por otro lado, generan una abundancia de entes diseñados, muchos de los cuales son competentes artesanos (constructores de nidos, tejedores de telarañas, etc.) y unos pocos incluso llegan a ser diseñadores y constructores inteligentes: nosotros.

Los procesos evolutivos introducen propósitos y razones del mismo modo que introducen la visión en color (y, por tanto, los colores): gradualmente. Si podemos entender cómo surgió nuestro mundo humano de razones a

partir de un mundo más simple en que no había razones, veremos que los propósitos y las razones son tan reales como los colores, tan reales como la vida misma. Los pensadores que insisten en que Darwin ha desterrado la teleología deberían añadir, por coherencia, que la ciencia también ha demostrado la irrealidad de los colores y de la propia vida. Todo lo que hay son átomos, y los átomos no tienen colores, ni están vivos. ¿Cómo puede ser que un gran conjunto de cosas sin color y sin vida forme algo con color y vivo? Se trata de una pregunta retórica, pero que (tarde o temprano) tiene que responderse, y puede ser respondida. Ahora me gustaría defender la afirmación de que hay razones para lo que hacen las proteínas, y para lo que hacen las bacterias, y para lo que hacen los árboles, y los animales, y nosotros (y, por supuesto, hay colores; y sí, Virginia, la vida existe).*

LOS DIFERENTES SENTIDOS DE «POR QUÉ»

Quizá la mejor manera de ver la realidad de las razones, de ver su ubicuidad en la naturaleza, es reflexionar sobre los diferentes sentidos de por qué. El término resulta equívoco; su ambigüedad queda clara si lo sustituimos por dos alternativas posibles: para qué y cómo es que.

«¿Por qué me das tu cámara?» pregunta para qué me la das, con qué intención, con qué propósito.

«¿Por qué flota el hielo?» pregunta cómo es que el hielo flota, cuál es la causa de que flote, qué es lo que le pasa al hielo cuando se forma que lo hace menos denso que el agua líquida.

La pregunta cómo es que nos demanda una narrativa del proceso que explique el fenómeno, sin plantear que sea para nada, sin plantear que haya ningún propósito. «¿Por qué es azul el cielo?» «¿Por qué los granos de arena en la playa están ordenados por tamaños?» «¿Por qué ha temblado el suelo?» «¿Por qué las tormentas vienen acompañadas de

granizo?» «¿Por qué este lodo se ha agrietado de esta forma concreta?» O también «¿Por qué ha fallado este álabe de la turbina?» Hay personas a quienes les gustaría considerar que la pregunta «por qué flota el hielo» es una invitación a hallar una razón del para qué (una razón divina, seguramente) de este hecho del mundo inanimado («Creo que Dios quería que los peces pudieran vivir bajo el hielo en invierno y si los lagos se congelaran de abajo hacia arriba, supondría un problema para los pobres peces»). Pero mientras tengamos una respuesta a la pregunta de cómo es que en términos de la física y la química, preguntar más sería casi paranoico.

Comparemos ahora estas cuatro preguntas:

¿Cuál es la razón de que los planetas sean esféricos?

¿Cuál es la razón de que los rodamientos sean esféricos?

¿Cuál es la razón de que los asteroides no sean esféricos?

¿Cuál es la razón de que los dados no sean esféricos?

El término razón me parece aceptable en las cuatro preguntas (¿a vosotros también os suena bien? ¿qué opináis?), pero las respuestas a las preguntas 1 y 3 no dan razones (no hay ninguna razón), dan causas, o narrativas de un proceso. Desgraciadamente, en algunos contextos, el término razón puede significar causa. Se puede responder a las preguntas 2 y 4 con narrativas de procesos, algo así como «bueno, los rodamientos se moldean en una especie de torno, que hace girar el metal… los dados se hacen en moldes cúbicos…», pero esto no son razones. A veces, la gente confunde las diferentes acepciones del término, como en una memorable conversación que tuve durante un debate con un fervoroso partidario del conductismo de Skinner, Lou Michaels, en la Western Michigan University en 1974. Yo había presentado mi artículo «Skinner Skinned» (en Brainstorms, 1978) y Michaels, en su refutación, presentó una pizca de ideología conductista especialmente audaz, a la que yo respondí «Pero ¿por qué dices esto, Lou?»; su respuesta inmediata fue «Porque me ha sido provechoso decirlo en todas las ocasiones pasadas». Yo estaba

pidiendo una razón, un para qué, pero como respuesta obtuve una narrativa del proceso, un cómo es que. Hay una diferencia, y el fracasado intento skinneriano de hacerla desaparecer debería prevenir a los científicos más positivistas de que, si intentan desterrar los para qué, pagarán un precio muy alto por comprender el mundo.

Las dos primeras frases de este libro han sido «¿Cómo es que hay mentes? ¿Y cómo es posible que estas mentes hagan esta pregunta y la respondan?» Son unas preguntas que piden una narrativa del proceso como respuesta, y es esto lo que voy a intentar ofrecer. Pero será una narrativa que también responderá las preguntas de cómo es que hay preguntas para qué y de para qué hay preguntas para qué.

LA EVOLUCIÓN DEL «POR QUÉ»: DEL «CÓMO ES QUE»

AL «PARA QUÉ»

La evolución por selección natural se inicia con un cómo es que y llega a un para qué. Empezamos con un mundo sin vida en el que no hay ningún tipo de razones, ningún tipo de propósitos, pero hay procesos: planetas en rotación, mareas, heladas y deshielos, erupciones volcánicas e incontables reacciones químicas. Algunos de estos procesos dan lugar a otros procesos que, a su vez, generan nuevos procesos, hasta que en algún punto (aunque no hay una frontera nítida) ya nos parece apropiado describir las razones por las que determinadas cosas están dispuestas de la manera en que lo están (¿Por qué nos parece adecuado describirlas? ¿Cómo es que llegamos a este estado mental? Paciencia; la respuesta no tardará en llegar).

Un rasgo fundamental de las interacciones humanas, y único de nuestra especie, es la actividad de pedir a los demás que se expliquen, que justifiquen sus decisiones y acciones y, luego, juzgar, respaldar o rebatir sus respuestas, en un ciclo recursivo de porqués. Los niños lo captan en seguida, llevando este juego al extremo y poniendo a prueba la paciencia de sus padres: «¿Por qué estás cortando este tablón de madera?» «Estoy

haciendo una puerta» «¿Por qué estás haciendo una puerta?» «Para poder cerrar la casa cuando salimos» «¿Por qué quieres cerrar la casa cuando salimos?» … «¿Por qué no queremos que los desconocidos cojan nuestras cosas?» … «¿Por qué tenemos cosas?» La soltura con la que todos nos embarcamos en este proceso de «control de razones» demuestra su importancia en nuestras vidas: nuestra capacidad de responder adecuadamente en esta actividad de control de razones es la base de la responsabilidad (Anscombe 1957). Aquellos que no se pueden explicar o que no se ven afectados por las razones ofrecidas por otros, aquellos que son sordos a las argumentaciones de otros, se consideran, con justificación, como personas con responsabilidad atenuada y la ley las trata de modo diferente.

No estamos continuamente enfrascados en esta actividad de pedir y evaluar las razones de los demás para sus acciones, pero tiene un papel importante para coordinar nuestras actividades, iniciar a los jóvenes en las actividades de los adultos y establecer normas con las cuales nos juzgamos unos a otros. Esta práctica es tan fundamental para nuestra forma de vida que a veces nos cuesta imaginar cómo otras especies pueden arreglárselas sin ella (como los delfines, los lobos o los chimpancés). ¿Cómo aprenden las crías su lugar en la comunidad, por ejemplo, sin que se les diga? ¿Cómo resuelven los elefantes las desavenencias acerca de cuándo y adónde desplazarse? Seguramente basta con sutiles señales instintivas de aprobación o desacuerdo; también debemos tener en cuenta que no hay ninguna otra especie que se enfrasque en comportamientos cooperativos con el grado de complejidad que hemos alcanzado los humanos.

Wilfrid Sellars, un filósofo de la Universidad de Pittsburgh, ha descrito esta actividad de intercambiar razones con los demás como la formación o constitución de «el espacio lógico de razones» (1962) y, con ello, ha empujado a toda una generación de filósofos de Pittsburgh, liderados por Robert Brandom y John Haugeland, a explorar este campo en profundidad. ¿Qué acciones están permitidas, y por qué? ¿Cómo entran en este espacio nuevas apreciaciones y cómo se gestionan las infracciones? El espacio de razones está limitado por normas, por el reconocimiento mutuo de cómo deberían ser las cosas, es decir, la forma correcta de participar en este juego de aportar razones. Así, siempre que hay razones, hay la posibilidad, y la necesidad, de algún tipo de justificación y la posibilidad de corrección si algo sale mal.

Esta «normativización» es la base de la ética: la capacidad de apreciar cómo debería funcionar el proceso de aportar razones es un prerrequisito para apreciar cómo debería ser la vida en sociedad. ¿Cómo y por qué surgió esta práctica y sus reglas? No ha existido siempre, pero ahora ciertamente sí existe. ¿Cómo es que existe, y para qué? Los filósofos de Pittsburgh no han abordado esta cuestión, preguntando cómo «llegó a ser así», de modo que tendremos que complementar su análisis con nuestras propias y prudentes especulaciones sobre la evolución del juego de aportar razones. Mostraré que, al ignorar esta cuestión, los filósofos de Pittsburgh han tendido a suprimir la distinción entre dos tipos diferentes de normas y sus respectivos modos de corrección, que denominaré normatividad social y normatividad instrumental. La primera, estudiada y celebrada en Pittsburgh, se refiere a las normas sociales que surgen con la práctica de la comunicación y la colaboración (de ahí que Haugeland [1998] hable de la «censura» de los miembros de la sociedad como fuerza correctora). La segunda se ocupa del control de calidad, de la eficiencia, de lo que podríamos denominar las normas de ingeniería que podemos ver en las fuerzas del mercado o en los fallos que se producen por causas naturales. Esto queda bellamente ilustrado con la distinción entre una buena obra y un buen utensilio; una buena obra puede estar torpemente ejecutada o incluso fracasar en su objetivo, mientras que un buen utensilio pude ser un efectivo instrumento de tortura o un arma funesta. Podemos observar el mismo contraste en los casos negativos, como la distinción entre travieso y bobo. La gente te puede castigar por ser travieso, de acuerdo con sus opiniones, pero la naturaleza te puede castigar ciegamente por ser bobo. Tal como veremos, necesitamos ambos tipos de normas para formar la perspectiva desde la cual podremos percibir las razones en la naturaleza.

La apreciación de las razones no evolucionó junto con las razones del modo en que la visión en color evolucionó junto con el color. El reconocimiento de las razones es un producto posterior y más avanzado de la evolución que las razones mismas.

Allí donde hay razones, siempre se puede invocar una norma implícita: se supone que las verdaderas razones son siempre buenas razones, razones que justifican el tema en cuestión (las preguntas cómo es que no implican ninguna demanda de justificación). Cuando aplicamos la ingeniería inversa a un artefacto recientemente descubierto, podemos preguntarnos, por ejemplo, por qué en una esquina hay ese botón tan evidente que no parece «hacer nada» (nada útil; proyecta una sombra cuando se ilumina y

modifica el centro de gravedad del artefacto, pero no tiene ningún propósito evidente). Hasta que se demuestre lo contrario, suponemos que el diseñador tuvo una razón, una buena razón para colocar ahí ese botón. Quizá en su momento era una buena razón, pero esa razón ya no es aplicable actualmente y los fabricantes lo han olvidado; se trataría, pues, de un botón vestigial, afuncional, que solo está presente a causa de la inercia del proceso de fabricación. Las mismas expectativas están detrás de la exploración de los seres vivos mediante ingeniería inversa; los biólogos aceptan a menudo hablar, informalmente, de lo que «la naturaleza pretende» o de lo que «tiene pensado la evolución» al «seleccionar» algún rasgo desconcertante de un ser vivo.[2] Sin duda la práctica de los biólogos es un descendiente directo de la ingeniería inversa aplicada a los objetos diseñados y elaborados por otros seres humanos, lo que, a su vez, desciende directamente de la institución social consistente en pedir y ofrecer razones para las actividades humanas. Esto podría significar que se trata de una práctica anticuada, un vestigio del pensamiento precientífico (y muchos biólogos suponen que así es), o bien podría significar que los biólogos han hallado una magnífica ampliación de la ingeniería inversa aplicable a la materia viva, usando los recursos de pensamiento con que la naturaleza nos ha dotado, para revelar pautas reales en el mundo que perfectamente pueden denominarse razones para la existencia de otras pautas reales. Para respaldar esta última afirmación, deberemos echar un vistazo a cómo se pudo poner en marcha la evolución.

CRECED Y MULTIPLICAOS

En mi libro La peligrosa idea de Darwin (1995) sostenía que la selección natural es un proceso algorítmico, un conjunto de algoritmos de ordenamiento formado, a su vez, por algoritmos de tanteo (también llamados de generación y prueba) que aprovechan la aleatoriedad (pseudoaleatoriedad, caos) en la fase de generación y algún tipo de control de calidad ciego en la fase de prueba, en el cual los ganadores siguen adelante con más descendencia que los otros. ¿Cómo se originó esta cascada de procesos generadores? Tal como he señalado en el último capítulo, el conjunto concreto de procesos que llevó al origen de la vida

aún nos es desconocido, pero podemos dispersar un poco la niebla que lo oculta señalando que, como es habitual, hay una gran variedad de procesos graduales de revisión que pueden dar el primer empujón.

El mundo prebiótico o abiótico no era un caos absoluto, un confeti aleatorio de átomos en movimiento. Concretamente, había ciclos a muchas escalas espaciotemporales: estaciones, día y noche, mareas, el ciclo del agua y miles de ciclos químicos observables a la escala atómica y molecular. Podemos considerar estos ciclos como los bucles FOR de los algoritmos informáticos, un grupo de acciones que vuelve al punto de partida tras realizar algo (acumular algo, desplazar algo u ordenar algo, por ejemplo) y luego se repite (y se repite, y se repite…); así, gradualmente, va cambiando las condiciones de su entorno y aumentando la probabilidad de que pase algo nuevo. Kessler y Werner presentaron un impresionante ejemplo abiótico de este proceso en la revista Science en 2003 (véase la figura 3.1).

FIGURA 3.1. Los círculos de piedras de Kessler y Werner. © Revista Science y Mark A. Kessler

Cualquier persona consideraría que estos círculos de piedras son una improbabilísima dispersión de piedras por el terreno; parecen claramente «artificiales» (y recuerdan a las elegantes esculturas al aire libre de Andy Goldsworthy), pero en realidad son el resultado natural de cientos de miles de ciclos sin rumbo de congelación y deshielo en la isla de Spitsbergen, en el Ártico. Desde hace siglos, los granjeros de Nueva Inglaterra saben que, cada invierno, las heladas hacen crecer una «nueva cosecha» de piedras en la superficie del suelo, piedras que tienen que retirarse antes de poder arar y plantar. Los típicos muros de piedras de Nueva Inglaterra que aún vemos hoy en día enmarcando los cultivos y, a veces, atravesar bosques producto de la reforestación de cultivos más antiguos, nunca han tenido la intención de servir de barrera a nada; en realidad no son muros, sino largos montones de rocas y pequeñas piedras transportadas por los granjeros hasta el borde de sus campos. Estos muros son sin duda el resultado de una acción humana deliberada con un propósito claro. Irónicamente, si los granjeros no hubieran retirado las piedras, a lo largo de muchos ciclos de heladas y deshielos, las piedras hubieran formado uno de los fenómenos de patrones en el suelo ilustrados aquí. Estos patrones no son siempre círculos; más a menudo se trata de polígonos, estructuras parecidas a laberintos y otras formaciones. Kessler y Werner ofrecen una explicación de este proceso mediante un modelo (un algoritmo) que genera estos diversos efectos de ordenamiento de las piedras variando parámetros como el tamaño de las piedras, la humedad, densidad y temperatura del suelo, la velocidad de congelación y la pendiente del terreno. Así, tenemos una idea bastante clara de cómo es que existe este fenómeno en los lugares en que se da; en cambio, cualquiera que haya visto estos círculos de piedras y haya llegado a la conclusión de que tenía que haber un artífice con una determinada intención, de que tenía que haber una respuesta a la pregunta para qué, estaría equivocado.

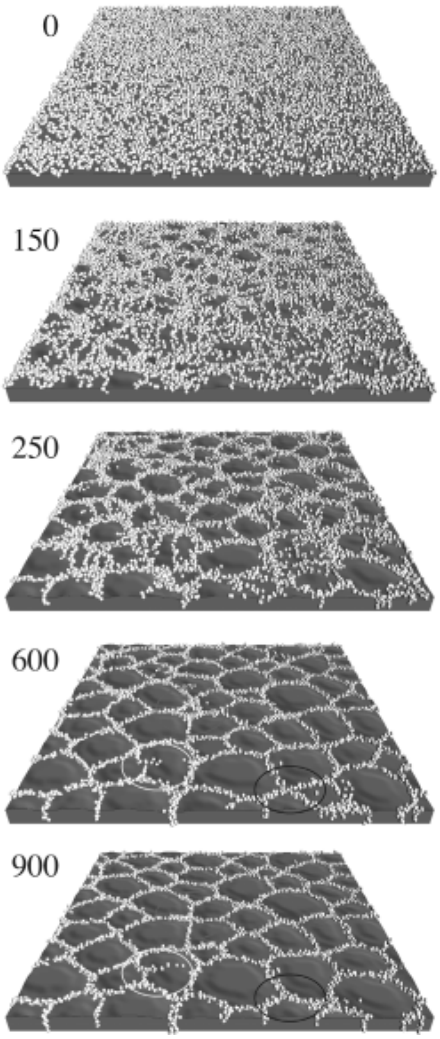


FIGURA 3.2. Algoritmo de ordenamiento de piedras de Stone y Kessler.
© Revista Science y Mark A. Kessler

En el mundo abiótico se producen muchos de estos ciclos, de modo simultáneo, pero no sincronizado, ruedas dentro de ruedas dentro de ruedas, con diferentes periodos de repetición, que «exploran» el espacio de posibilidades químicas. Sería algo así como una variación del procesamiento en paralelo, no muy diferente de la fabricación en masa de la industria, que hace piezas diferentes en diferentes lugares a ritmos de producción también diferentes y, finalmente, lo ensambla todo para formar el conjunto; la diferencia es que esta producción abiótica en masa no tiene ningún tipo de planificación ni propósito. En el mundo abiótico no hay re-producción diferencial, pero sí que se obtienen diversos tipos de persistencia diferencial: algunas combinaciones temporales de partes

están presentes más tiempo que otras, con lo que disponen de algo más de tiempo para adquirir posibles mejoras y ajustes. En cierto modo, los ricos se hacen más ricos, aunque no puedan legar esta riqueza a sus descendientes. Esta persistencia diferencial, de algún modo, tiene que convertirse gradualmente en reproducción diferencial. Los algoritmos protodarwinianos de supervivencia diferencial de combinaciones químicas pueden dar lugar a ciclos de reacciones autocatalíticas que, a su vez, dan lugar a una replicación diferencial como un caso particular, muy especial, de persistencia diferencial, de un tipo particularmente explosivo que multiplica su ventaja… ¡multiplicándose a sí mismo! Esta replicación diferencial genera muchos elementos persistentes que son duplicados con ligeros cambios respecto a los anteriores y que, así, pueden «explorar» muchos más rincones ligeramente diferentes del mundo que cualquier otro elemento.

Dicen que «un diamante es para siempre», pero es un dicho algo exagerado. Un diamante es extraordinariamente persistente, mucho más que sus competidores habituales, pero su persistencia está modelada por su avance lineal a través del tiempo: el diamante del martes es igual a su padre, el diamante del lunes, y así sucesivamente. No se multiplica nunca, pero puede acumular cambios, desgaste y deterioro, una capa de suciedad que se endurece, etc., cambios que lo pueden hacer más persistente o menos. Al igual que otros objetos duraderos, se ve afectado por muchos ciclos, muchos bucles FOR en los que está implicado de un modo u otro. Normalmente estos efectos no se acumulan durante mucho tiempo, sino que son barridos por otros efectos posteriores; a veces, sin embargo, se forma algo parecido a una barrera: una pared o una membrana que proporciona una protección adicional.

En el mundo del software, dos fenómenos bien conocidos son la serendipia y su opuesto, el machacado o sobreescritura accidental. Ambos consisten en el choque fortuito de dos procesos inconexos, pero mientras que en el primero el resultado es afortunado, en el segundo es destructivo. Los muros o membranas que, por la razón que sea, tienden a evitar fenómenos de machacado serán especialmente persistentes y permitirán que los bucles FOR internos actúen sin interferencias. Así, vemos la necesidad de membranas para albergar el conjunto de ciclos químicos (el ciclo de Krebs y miles más) que permiten el surgimiento de la vida (una referencia excelente para este punto de vista algorítmico de los ciclos químicos en las células es el libro Wetware de Dennis Bray). Incluso las células bacterianas más simples tienen algo parecido a un sistema

nervioso formado por redes químicas de una eficiencia y elegancia extraordinarias. Ahora bien, ¿cómo pudo llegar a surgir la combinación adecuada de membranas y bucles en el mundo prebiótico? Algunos pueden decir «¡Imposible! ¡Ni en un millón de años!». Vale, de acuerdo, pero ¿qué tal cien millones de años? Solo tiene que suceder una vez para que el motor de la reproducción se ponga en marcha.

Imaginemos que estamos en los primeros momentos de este proceso por el cual la persistencia se convierte gradualmente en multiplicación, observamos la proliferación de algún tipo de elementos allí donde antes no había ninguno y nos preguntamos «¿Por qué estamos observando aquí estos objetos tan improbables?» ¡La pregunta es ambigua! Es ambigua porque, ahora, como respuesta hay tanto una narrativa del proceso (un cómo es que) y una justificación (un para qué). Estamos ante una situación en la que hay determinadas estructuras químicas, mientras que otras alternativas químicas posibles están ausentes, y estamos buscando cosas que sean mejores persistiendo en las circunstancias locales que sus posibles alternativas contemporáneas. Antes de poder tener reproductores competentes, debemos tener «persistores» competentes, estructuras con la suficiente estabilidad para durar lo bastante y así poder adquirir revisiones. Sin duda, no parece una habilidad muy impresionante, pero es lo que basta para iniciar la historia darwiniana: algo que sea solo más o menos competente, nada muy sofisticado. Lo que estamos presenciando es un descarte automático (algorítmico) de lo no funcional, desplazado por lo funcional; y cuando lleguemos a observar una bacteria reproductora, habrá una abundancia de virtuosismo funcional. Dicho de otro modo, hay razones por las que las diferentes partes están conformadas y dispuestas del modo en que lo están. Podemos aplicar ingeniería inversa a cualquier ente que se reproduce, determinar lo que tiene de bueno y lo que tiene de malo y explicar por qué es bueno o malo. He aquí el nacimiento de las razones, y es agradable señalar que se trata de un caso de lo que Glenn Adelson ha llamado acertadamente darwinismo del darwinismo (Godfrey-Smith 2009): observamos el surgimiento gradual de las especies de razones a partir de las especies de meras causas, de los para qué a partir de los cómo es que, sin ninguna frontera fundamental trazada entre ellos. Del mismo modo que no hay ningún Primer Mamífero (el primer mamífero que no tuvo un mamífero como madre), tampoco hay ninguna Primera Razón, el primer rasgo de la biosfera que ayudó a que algo existiera porque lo hacía mejor en existir que su «competencia».

La selección natural es un buscador automático de razones, que «descubre», «respalda» y «se centra en» las razones a lo largo de muchas generaciones. Lo pongo entre comillas para recordar que la selección natural no tiene una mente, no tiene razones por sí misma, pero sin embargo es competente en esta tarea de perfeccionar los diseños. Asegurémonos de que sabemos cómo eliminar las comillas. Consideremos una población de escarabajos con mucha variación; algunos son buenos multiplicándose, la mayoría, no tanto. Ahora consideremos la minoría que (habitualmente) es buena en esta tarea y preguntémonos sobre cada uno: ¿por qué lo ha hecho mejor que la media? Nuestra pregunta es ambigua; se puede interpretar como un cómo es que o como un para qué. En muchos casos, en la mayoría, de hecho, la respuesta es que no hay ninguna razón, se trata simplemente de pura suerte, buena o mala; en estos casos solo tenemos una respuesta cómo es que a nuestra pregunta. Pero si hay un subconjunto, quizá muy pequeño, de casos en que hay una respuesta a la pregunta para qué, una diferencia que resulta que marca la diferencia, todos estos casos tienen en común el germen de una razón, una protorrazón, si queréis llamarla así. La narrativa del proceso explica cómo llegó a ser y, además, señala por qué estos son mejores que aquellos, por qué ganaron en la competición. El lema del torneo de la evolución es «que gane el mejor» y los ganadores, por ser los mejores, llevan consigo la justificación de sus mejoras. En cada generación, en cada estirpe, solamente algunos competidores logran reproducirse, y cada descendiente de la siguiente generación o bien es simplemente afortunado o es afortunado de estar bien dotado de alguna manera. El último grupo resultó seleccionado («por alguna causa», podríais decir, aunque sería mejor «por alguna razón»). Este proceso explica la acumulación de funciones mediante un proceso que rastrea a ciegas las razones y crea cosas que tienen propósitos, pero no necesitan saberlos. El principio de mínimo conocimiento (need-

to-know principle), tan habitual en las novelas de espías, también domina la biosfera: un organismo no necesita saber las razones por las cuales los dones que hereda son ventajosos para él, y la selección natural tampoco necesita saber qué es lo que está haciendo.

Darwin entendió esta idea:

En cierto modo, el término selección natural es una mala opción, ya que parece implicar una elección consciente, pero esta idea se desestimará después de familiarizarse un poco. Nadie reprocha a los químicos que hablen de afinidades electivas y está claro que un ácido no tiene más capacidad de elección al combinarse con una base que la que tienen las condiciones de la vida al determinar si una nueva forma se selecciona o conserva […] Por comodidad, a veces hablo de la selección natural como un poder inteligente, de la misma forma que los astrónomos dicen que la atracción de la gravedad determina el movimiento de los planetas. […] A menudo, también he personificado la naturaleza, pues me ha resultado difícil evitar esta ambigüedad; pero con naturaleza solo me refiero a la acción conjunta y el resultado de muchas leyes naturales; y con leyes, a la sucesión establecida de sucesos. (1868)

Así pues, había razones mucho antes de que hubiera algo que pudiera representar estas razones (nosotros). A las razones que la evolución rastrea, las he denominado motivaciones flotantes, un término que, según parece, ha puesto algo nerviosos a algunos pensadores, que temen que esté conjurando algún tipo de fantasmas. En absoluto, las motivaciones flotantes no son más fantasmales ni más problemáticas que los números o que los centros de gravedad. Los cubos ya tenían ocho vértices antes de que los humanos inventásemos modos de formular la aritmética, y los asteroides ya tenían centros de gravedad antes de que los físicos ni tan siquiera hubieran concebido la idea y los hubieran calculado. Las razones existían mucho antes de que hubiera razonadores. Algunas personas creen que esta forma de pensar resulta inquietante y, seguramente, irracional, pero no por ello voy a rendirme; lo que espero poder hacer aquí es tranquilizarles y convencerles de que deberíamos considerarnos afortunados de poder hablar de las razones descubiertas por la evolución antes de que los investigadores humanos (u otras mentes) las expresaran y las representaran.[3]

Examinemos ahora las dos construcciones, increíblemente parecidas, de las figuras 3.3 y 3.4. El castillo de las termitas y la Sagrada Familia de Gaudí tienen una forma muy similar, pero son totalmente diferentes en su origen y su construcción. Hay razones para las estructuras y las formas del castillo de las termitas, pero ninguna de las termitas que lo ha construido es capaz de representar estas razones; no hay ninguna Termita Arquitecto que haya planificado la estructura y cada termita individual no

tiene la más mínima idea de por qué construyen de la forma en que lo hacen. He aquí un ejemplo de competencia sin comprensión, sobre la que hablaremos más adelante. Sí que hay razones para las estructuras y las formas de la obra maestra de Gaudí, pero, en su mayor parte, son las razones de Gaudí. Gaudí tenía razones para las formas que decidió crear; también hay razones para las formas creadas por las termitas, pero las termitas no poseen estas razones. Hay razones por las que los árboles abren las ramas en sus copas, pero de ningún modo son las razones de los árboles. Las esponjas hacen cosas por razones, las bacterias hacen cosas por razones, incluso los virus hacen cosas por razones; pero ninguno de ellos posee estas razones, no necesitan poseerlas.

FIGURA 3.3. Castillo de termitas australianas

FIGURA 3.4. Gaudí: La Sagrada Familia

¿Somos nosotros, humanos, los únicos seres que pueden representar razones? Es una pregunta muy importante, pero aplazaré la respuesta hasta que dispongamos de un panorama más amplio para el cambio de perspectiva que propongo. Hasta el momento, lo que espero haber demostrado es que Darwin no acabó con la teleología, la naturalizó; ahora bien, esta conclusión no está tan aceptada como debiera y a algunos científicos les da una cierta aprensión y les hace actuar con excesiva cautela, evitando cualquier discusión sobre diseño y sobre razones. La práctica humana de ofrecer razones crea el espacio de razones, que queda limitado por las normas, tanto sociales y éticas como instrumentales (la diferencia entre ser travieso y ser bobo). La ingeniería inversa en biología desciende de este proceso de ofrecer y juzgar razones.

La evolución del cómo es que al para qué se puede observar en el modo en que interpretamos el surgimiento gradual de los seres vivos mediante una cascada de ciclos prebióticos. Las motivaciones flotantes surgen como las razones por las cuales existen determinados rasgos; no presuponen diseñadores inteligentes, aunque los diseños que emergen sean increíblemente buenos. Así, por ejemplo, hay razones por las que las colonias de termitas presentan las características que presentan, pero las termitas, a diferencia de Gaudí, no poseen ni representan razones y sus excelentes diseños no son el producto de ningún diseñador inteligente.

---

* El autor se refiere a la frase «Yes, Virginia, there is a Santa Claus» (‘sí, Virginia, Santa Claus existe’), de un famoso editorial publicado en el The Sun de Nueva York el 21 de septiembre de 1897 y que se ha convertido en un elemento muy popular de la tradición navideña de Estados Unidos. (N. del t.)

4

DOS EXTRAÑAS INVERSIONES DEL RAZONAMIENTO

EL HECHIZO QUE ROMPIERON DARWIN Y TURING

Antes de Darwin, el mundo se sostenía gracias a la tradición, no gracias a la ciencia. Todas las cosas del universo, desde la más excelsa («el hombre») a la más humilde (la hormiga, el guijarro, la gota de lluvia) eran creaciones de algo aún más excelso: Dios, un creador inteligente, omnipotente y omnisciente, con un extraordinario parecido con la segunda cosa más excelsa. Llamemos a esto la «teoría de la creación por goteo»; Darwin la sustituyó por la «teoría de la creación por burbujeo». Robert MacKenzie Beverley, uno de los críticos de Darwin del siglo XIX,[1] lo expresó con claridad:

En la teoría con la que tenemos que tratar, el artífice es la Ignorancia Absoluta; así, podemos enunciar como principio fundamental de todo el sistema, que, PARA HACER UNA MÁQUINA PERFECTA Y BELLA, NO ES NECESARIO SABER CÓMO HACERLA. Con un examen cuidadoso se verá que esta proposición expresa, de modo condensado, el significado básico de la Teoría y plantea en pocas palabras toda la intención del Sr. Darwin, quien, mediante una extraña inversión del razonamiento, parece creer que la Ignorancia Absoluta está plenamente capacitada para sustituir a la Sabiduría Absoluta en los logros de la habilidad creadora. (Beverley 1868.)

Sin duda, esto fue una «extraña inversión del razonamiento» y la incredulidad expresada por Beverley aún perdura en una gran parte de la población del siglo XXI.

Cuando consideramos la teoría de la creación por burbujeo de Darwin, podemos imaginar una metáfora en que todo el trabajo de diseño creativo va encaramándose en lo que yo denomino espacio de diseño. Tiene que empezar con los primeros burdos replicadores que vimos en el capítulo 3 y, poco a poco, ir subiendo, oleada tras oleada de selección natural, hasta la vida multicelular en todas sus formas. Pero ¿de verdad puede este proceso haber dado lugar a todas las maravillas que observamos en la biosfera? Desde la época de Darwin, los escépticos han intentado demostrar que resulta imposible acercarse a alguna de estas maravillas siguiendo esta ruta difícil y carente de inteligencia; han buscado algo vivo pero «inevolucionable». Suelo usar la denominación gancho celeste para referirme a este enfoque, como un instrumento imaginario que puedes colgar en el cielo para sostener una polea o lo que sea que quieras alzar (Dennett 1995). Un gancho celeste flota muy alto en el espacio de diseño, sin estar sostenido por sus progenitores; es el resultado directo de un acto especial de creación inteligente. Sin embargo, una y otra vez, estos escépticos han descubierto que no hay ningún gancho celeste milagroso, sino una grúa extraordinaria, una innovación para nada milagrosa en el espacio de diseño que permite una exploración aún más eficiente de las posibilidades de diseño y un izamiento aún más potente en el espacio de diseño. La endosimbiosis es una de estas grúas: alzó las células simples y aisladas hasta un mundo de enorme complejidad, desde el cual pudo despegar la vida multicelular. El sexo es otra de estas grúas: permitió que el patrimonio génico se mezclara y, así, pudiera ser explorado de forma mucho más efectiva por los procesos a ciegas de prueba y error de la selección natural. El lenguaje y la cultura son grúas, novedades que evolucionaron y que abrieron amplísimos espacios de posibilidades para que diseñadores aún más inteligentes (pero no milagrosamente inteligentes) los pudieran explorar. Sin añadir el lenguaje y la cultura al arsenal de recursos de I+D disponibles para la evolución, nunca habría habido plantas de tabaco bioluminiscentes que contienen genes de luciérnagas; no se trata de un milagro, son un fruto del árbol de la vida al mismo nivel que las telarañas o las presas de los castores, pero la probabilidad de que hubieran surgido sin la asistencia de Homo sapiens y de nuestros recursos culturales es nula.

Cuanto más sabemos acerca de la nanomaquinaria de la vida que permite todo esto, más podemos apreciar otra extraña inversión del razonamiento, conseguida casi un siglo después por otro brillante sabio británico: Alan Turing. Usando el mismo lenguaje que Beverley, podemos expresar la extraña inversión de Turing así:

PARA SER UNA MÁQUINA DE CALCULAR PERFECTA Y BELLA, NO ES NECESARIO SABER QUÉ ES LA ARITMÉTICA.

Antes de Turing ya había computadoras, a cientos o a miles, usadas para efectuar cálculos científicos y de ingeniería. Las computadoras eran personas, no máquinas; muchas eran mujeres y muchas eran licenciadas en matemáticas. Eran seres humanos que sabían qué era la aritmética, pero Turing tuvo una brillante intuición: ¡no necesitaban saberlo! Tal como señaló, «el comportamiento de la computadora en cualquier momento está determinado por los símbolos que observa y por su "estado mental" en ese momento» (Turing 1936, 5). Ese «estado mental» era un conjunto muy simple de instrucciones condicionales IF-THEN acerca de qué hacer y a qué «estado mental» pasar a continuación (y así sucesivamente hasta encontrar una instrucción STOP). Turing demostró que era posible diseñar máquinas absolutamente ignorantes pero que podían hacer cálculos aritméticos perfectos, siguiendo unas «instrucciones» que se podían introducir de forma mecánica. Es más, demostró que si las instrucciones incluían ramificaciones condicionales (instrucciones IF-THEN, como «si te encuentras un 0, sustitúyelo por un 1 y desplázate a la izquierda, y si te encuentras un 1, déjalo como está y desplázate a la derecha; luego cambia tu estado a n»), las máquinas podían seguir indefinidamente rutas complejas determinadas por las instrucciones, lo que les aportaría una capacidad digna de destacar: podrían hacer cualquier cosa computacional. Dicho de otro modo, una computadora digital programable es una máquina de Turing universal, que puede imitar a cualquier computadora especializada siguiendo un conjunto de instrucciones que pongan en práctica, como software, esa computadora especializada (no necesitas volver a cablear tu teléfono para que pueda hacer nuevas tareas; basta con bajar una aplicación y convertir el teléfono en un planetario, un traductor, una calculadora o un

revisor ortográfico).[2] Turing abrió la puerta a un enorme espacio de diseño relativo al procesamiento de información y también previó que en ese espacio de diseño había un camino transitable entre la ignorancia absoluta y la inteligencia artificial, un largo trecho de escalones ascendentes.

Muchas personas no pueden soportar la extraña inversión de Darwin; son los que denominamos creacionistas, que continúan buscando los ganchos celestes, rasgos de la biosfera «irreductiblemente complejos» (Behe 1996) que no puedan haber surgido a través de procesos darwinianos. Pero incluso más gente es incapaz de soportar la extraña inversión de Turing, y por razones sorprendentemente parecidas. Quieren creer que las maravillas de la mente no pueden alcanzarse por meros procesos materiales, que las mentes son, si no milagrosas en su sentido estricto, sí misteriosas de un modo que desafía a las ciencias naturales. No desean que se cicatrice la herida cartesiana.

¿Y por qué no? Ya hemos señalado algunos de sus motivos menos edificantes: miedo, orgullo, la inapropiada atracción de los misterios sin resolver. Ahora encontramos otra razón (¿es un cómo es que o un para qué?): tanto Darwin como Turing afirman haber descubierto algo profundamente inquietante para la mente humana: competencia sin comprensión. Beverley expresó su indignación con entusiasmo: ¡la idea misma de habilidades creadoras sin inteligencia! Tengamos en cuenta cuán impactante es esta posibilidad y cómo se opone a una idea bien consagrada en nuestras políticas y prácticas educativas: la comprensión como la (mejor) fuente de la competencia. Enviamos a nuestros hijos a la universidad para que puedan comprender cómo funciona el mundo y que eso les sirva durante toda su vida, obteniendo competencias cuando lo necesiten a partir del valioso acopio de conocimientos que les hemos inculcado. ¿Por qué actualmente despreciamos el aprendizaje de memoria? Porque hemos visto (¿no es cierto?) que hacer que los niños comprendan un tema o un método es la manera (¿la única o solo la mejor?) de hacerlos competentes en ese tema o método. Despreciamos la memorización estúpida que se limita a rellenar los agujeros en una plantilla sin pararse a pensar por qué; nos burlamos de la idea de que colorear un dibujo por números pueda ser la manera de formar artistas creativos. En resumen, nuestro lema podría ser

Si consigues que comprendan, sus competencias vendrán después.

Subrayemos que aquí hay algo más que un poco de ideología. Estamos bastante familiarizados con algunos fracasos estrepitosos de este principio sagrado, como las «nuevas matemáticas», que, sin éxito, han tratado de enseñar a los niños teoría de conjuntos y otros conceptos abstractos en lugar de instruirlos en sumas y restas, las tablas de multiplicar, fracciones y algoritmos simples como la división larga o contar de dos en dos, de cinco en cinco, etc.

Los ejércitos son una de las instituciones educativas más eficaces del mundo, que convierten a alumnos normales de secundaria en fiables mecánicos de reactores, operadores de radar, pilotos y muchos otros tipos de especialistas técnicos gracias a una intensa instrucción. Poco a poco, surge una valiosa diversidad de comprensión a partir de las competencias inculcadas durante estas prácticas, con lo que tenemos una buena prueba empírica de que la competencia no siempre depende de la comprensión e incluso, a veces, es una condición previa para la comprensión. Lo que Darwin y Turing concibieron fue la versión más radical de esta idea: toda la comprensión y la genialidad de este mundo surge, en último término, de competencias sin comprensión, que se van combinando a lo largo del tiempo y forman sistemas cada vez más competentes (y, por ende, también más capaces de comprender). Esto es, en efecto, una extraña inversión, que derroca el punto de vista predarwiniano de la creación como un asunto de «al principio, la mente» y lo sustituye por una perspectiva de «al final, la mente» acerca de nuestra propia evolución, como diseñadores inteligentes que somos.

Nuestro escepticismo sobre la competencia sin comprensión tiene causas, no razones. No es que parezca lógico que no pueda haber competencia sin comprensión, simplemente parece correcto; y parece correcto porque nuestras mentes han sido conformadas para pensar de este modo. Le tocó a Darwin romper el hechizo lanzado por esta forma de pensar; Turing, poco después, lo volvió a romper para abarcar la idea novedosa de que podríamos invertir el orden tradicional y erigir comprensión a partir de una cascada de competencias, de forma parecida a cómo la evolución por selección natural erige arreglos internos,

órganos e instintos cada vez más brillantes sin tener que comprender qué es lo que está haciendo.

Ahora bien, hay una gran diferencia entre la extraña inversión de Darwin y la de Turing. Darwin demostró que cascadas de procesos sin ninguna inteligencia pueden crear diseños espléndidos, pero las cascadas de procesos de Turing son el producto de un diseñador sin duda inteligente: Alan Turing. Podríamos decir que mientras que Darwin descubrió la evolución por selección natural, Turing inventó la computadora (y subrayo la diferencia entre descubrió e inventó). Muchas personas sostienen que, para que pudiera darse la evolución por selección natural, un dios inteligente tuvo que disponer toda la situación de partida, y Turing parece representar exactamente este papel al establecer la idea esencial de una computadora (material, sin vida y sin comprensión) que puede ser el escenario en que aparezca la comprensión mediante un proceso no muy diferente de la evolución, un cúmulo de mejoras de diseño confeccionadas a partir de los ladrillos básicos de la computación. ¿Acaso el papel de Turing no se opone, en lugar de complementar, al alcance de la extraña inversión de Darwin? La respuesta es que no, y justificar esta respuesta es una de las tareas más relevantes del resto de este libro. La explicación breve es que Turing mismo no es más que una ramita del árbol de la vida y sus creaciones, concretas y abstractas, son el producto indirecto de los procesos darwinianos ciegos del mismo modo que lo son las telarañas o las presas de los castores; no hay ninguna discontinuidad radical, ninguna necesidad de un gancho celeste, para pasar de las arañas y los castores a Turing y las máquinas de Turing. Sin duda, hay un largo trecho a recorrer entre ambos puntos, porque la forma de Turing de hacer cosas es notablemente diferente de la de la araña y de la del castor, y necesitamos una buena descripción evolutiva de esta diferencia; porque ahora surge un problema: si la competencia sin comprensión es tan y tan fecunda (capaz de diseñar un ruiseñor, por ejemplo), ¿por qué necesitamos comprensión? (capaz de diseñar odas a ruiseñores** y computadoras) ¿Por qué y cómo apareció en escena la comprensión al estilo de los humanos? Para empezar, vale la pena subrayar con más fuerza este contraste.

Mientras que las termitas son un ejemplo asombroso de competencia sin comprensión, constructoras de hogares resistentes, seguros y con aire acondicionado sin necesidad de planos ni capataces (la termita reina es más un tesoro que proteger que un maestro de obra), Antoni Gaudí es un

ejemplo casi perfecto de diseñador inteligente, un jefe omnipotente, equipado desde el principio con planos, dibujos y escritos llenos de explicaciones justificadas con pasión. Su gran basílica de Barcelona es un ejemplo de creación «descendente» o «de arriba abajo» difícil de superar, aunque la computadora original de Turing, la Pilot ACE (que se puede ver actualmente en el Museo de la Ciencia de Londres), bien podría arrebatarle la primera posición. Se trata de una de las primeras computadoras realmente útiles, que se puso en funcionamiento en 1950 en el National Physical Laboratory de Inglaterra, y compite con la Sagrada Familia en originalidad y complejidad… ¡y en coste! Ambos creadores tuvieron que convencer a mecenas para que patrocinaran sus ambiciosos diseños y elaboraron detallados diagramas y redactaron abundantes explicaciones. En definitiva, en ambos casos, la realidad final dependía de la existencia de representaciones previas, en la mente de un genio, del propósito del diseño y, en consecuencia, de la raison d'être de cada una de sus partes.[3] En el momento práctico de la construcción real del objeto, los trabajadores tenían una comprensión mucho más limitada, una apreciación mínima de la finalidad de su tarea concreta. La comprensión estaba distribuida: Gaudí no necesitaba conocer al detalle cómo mezclar el mortero o cómo tallar la piedra, tal como sí lo sabían los albañiles, y Turing no tenía que ser un experto con el soldador ni con las técnicas de fabricación de tubos de vacío. La distribución de la pericia y del conocimiento es un rasgo distintivo de los proyectos creativos humanos y, sin duda, es fundamental para los artefactos actua-

les de alta tecnología, aunque no para los más sencillos. Un artífice solitario puede fabricar una lanza o incluso una canoa, un carro de madera o una choza de paja, y entender todos y cada uno de los aspectos del diseño y la construcción, pero esto sería imposible para fabricar una radio, un automóvil o una central nuclear.

Una mirada más atenta a unos pocos ejemplos de artefactos humanos y a la tecnología que hemos inventado para confeccionarlos nos servirá para identificar los diversos apeaderos que hay en el camino entre las bacterias sin rumbo y Bach. Antes, sin embargo, tenemos que presentar un término que se originó en la filosofía y que ha ampliado su campo para englobar diversas iniciativas científicas y de ingeniería.

ONTOLOGÍA Y LA IMAGEN MANIFIESTA

El término ontología procede de la palabra griega que significa cosa. En filosofía, se refiere al conjunto de «cosas» que una persona cree que existen, o al conjunto de cosas definidas, o asumidas, por una teoría. ¿Qué hay en tu ontología? ¿Crees en fantasmas? Entonces en tu ontología hay fantasmas, además de mesas y sillas y canciones y vacaciones y nieve y todo lo demás. Se ha demostrado especialmente útil ampliar el significado del término y usarlo también para el conjunto de cosas que un animal puede reconocer y ante las cuales puede actuar adecuadamente (con independencia de si podemos decir que un animal tiene creencias), así como, más recientemente, para el conjunto de cosas de las que un programa informático tiene que ocuparse para realizar su tarea (con independencia de si podemos decir que un programa tiene creencias). Las vacaciones no pertenecen a la ontología de un oso polar, pero la nieve sí, y las focas también; en cambio, la nieve seguramente no forma parte de la ontología de un manatí, pero las hélices de un motor fueraborda quizá sí, junto con las algas, los peces y otros manatís. El GPS de vuestro coche se ocupa de calles de sentido único, de giros a derecha e izquierda, de límites de velocidad y de la velocidad del coche en cada instante (si no es cero, puede que no os deje introducir un nuevo destino), pero su ontología, además, también incluye diversos satélites y las señales emitidas y recibidas de estos satélites, de las que no os informa, pero que necesita para realizar su tarea.

La ontología de un GPS fue sabiamente creada por sus programadores; seguramente, el proceso de I+D implicó muchas pruebas y errores a medida que se ensayaban diferentes diseños que, uno tras otro, resultaban deficientes, hasta encontrar el más adecuado. La ontología de un oso polar o de un manatí fue diseñada por alguna combinación de evolución genética y experiencia individual, difícil de discernir. Los manatís deben de tener las algas en su ontología de forma parecida a como los bebés humanos tienen los pezones en la suya: de un modo instintivo, conformado genéticamente a lo largo de las eras. Cualquier manatí que tenga las hélices de un fueraborda en su ontología, lo ha obtenido a través de la experiencia. Los humanos tenemos ontologías variadísimas; algunos creen en brujas, otros creen en electrones, otros, en resonancias mórficas y en abominables hombres de las nieves. Sin

embargo, hay un grandísimo acervo común de ontología, compartido por todos los seres humanos normales a partir de una edad bastante temprana (a los seis años ya casi se dispone de toda esta ontología).

Esta ontología común fue bautizada acertadamente como imagen manifiesta por Wilfrid Sellars (1962). Pensemos en el mundo en que vivimos, lleno de otras personas, de plantas, de animales, de muebles, de casas, de coches… y de colores, de arco iris, de puestas de sol, de voces, de peinados, de goles, de euros, de problemas, de oportunidades, de errores y de montones de cosas más. Estas son las incontables cosas que podemos identificar, señalar, amar u odiar y, en muchos casos, también manipular o incluso crear (no podemos crear puestas de sol, pero si disponemos adecuadamente algo de agua, con un poco de ingenio podemos crear un arco iris); son las cosas que usamos en los quehaceres cotidianos para fijar nuestras interacciones y conversaciones. De forma aproximada, a cada sustantivo de nuestra charla diaria le corresponde un tipo de cosa; es en este sentido en que la «imagen» es «manifiesta»: es obvia para todo el mundo, y todo el mundo sabe que es obvia para todos, y esto último también lo sabe todo el mundo. Acompaña a nuestra lengua materna; es el mundo según nosotros.[4] Sellars oponía esta imagen manifiesta a la imagen científica, poblada por moléculas, átomos, electrones, gravedad, quarks y quién sabe qué más (¿Energía oscura? ¿Cuerdas? ¿Branas?). Los científicos también pasan la mayor parte de sus vidas imaginando lo que está pasando en términos de la imagen manifiesta («dame el lápiz, por favor» es un pequeño fragmento de comunicación que depende de la imagen manifiesta, con las personas y sus necesidades y pasiones; con sus capacidades de oír, ver, comprender y actuar; con los rasgos distintivos de un lápiz, su tamaño y peso y su uso; y con muchas otras cosas. Fabricar un robot que pueda comprender y consentir una petición así no es en absoluto trivial, a no ser que hagas un robot que pueda «comprender» solamente esta frase y unas pocas más).

La imagen científica es algo que se tiene que aprender en la escuela, y la mayoría de personas solamente adquieren un conocimiento muy somero de ella. Hoy en día, estas dos versiones del mundo son muy diferentes, casi como dos especies independientes, pero hubo una época en que estaban mezcladas y entretejidas en un único mundo ancestral de «lo que todo el mundo sabe», que incluía toda la fauna y la flora local, las armas y los utensilios, las viviendas y los papeles sociales, pero también

duendes, dioses, miasmas y conjuros que podían maldecir tu vida o asegurarte una buena caza. Poco a poco, nuestros antepasados aprendieron a expulsar de sus ontologías las cosas innecesarias y a crear categorías para las nuevas; así echaron por la ventana a brujas, sirenas y gnomos y dieron la bienvenida a átomos, moléculas y gérmenes. Los primeros pensadores protocientíficos, como Aristóteles, Lucrecio y, mucho después, Galileo, llevaron a cabo sus investigaciones sin hacer una distinción nítida entre la ontología cotidiana (la imagen manifiesta) y la ontología científica, pero aun así, hicieron audaces propuestas de nuevos tipos de cosas, las más convincentes de las cuales tuvieron éxito y se mantuvieron. Una de las tareas más importantes de la ciencia moderna ha sido enmendar algunos de sus errores más tentadores y, a la vez, crear la ontología de la imagen científica.

A diferencia del término ontología, imagen manifiesta e imagen científica no han pasado de la filosofía a otros campos, pero hago lo que puedo para exportar estas denominaciones, pues desde hace mucho tiempo, me parecen la mejor manera de aclarar la relación entre «nuestro» mundo y el mundo de la ciencia. ¿De dónde procede la imagen manifiesta precientífica? Sellars se centró en la imagen manifiesta de los seres humanos o de las sociedades humanas; ¿deberíamos ampliar el concepto a otras especies? En un sentido amplio, tienen ontologías; ¿tienen también imágenes manifiestas? Y si es así, ¿son muy diferentes de las nuestras? Estas preguntas son relevantes para nuestra indagación, porque para comprender la grandeza de la extraña inversión de Darwin, tenemos que entender qué es lo que Darwin estaba invirtiendo y cómo apareció.

EL ASCENSOR AUTOMATIZADO

Lo mejor es empezar por lo más simple, con un ejemplo que no tiene nada que ver con la conciencia, ni tan solo con la vida: el dispositivo electrónico que controla un ascensor automatizado. Cuando yo era joven aún existían ascensores operados por personas, los ascensoristas, cuyo trabajo era subir y bajar todo el día con el ascensor, parando en las plantas adecuadas para dejar y recoger pasajeros. En los primeros años,

accionaban una curiosa palanca que se desplazaba en sentido horario o antihorario para que el ascensor subiera o bajara, y debían tener la habilidad suficiente para detener el ascensor justo a la altura correcta. A menudo, la gente tenía que subir o bajar unos centímetros para entrar o salir del ascensor, y los ascensoristas siempre avisaban de ello. Tenían muchas normas acerca de qué decir en cada momento, a qué plantas ir primero, cómo abrir las puertas, etc. Su formación consistía en memorizar estas reglas hasta que se convertían en algo instintivo. Las reglas, por su parte, se habían ido perfilando a lo largo de los años en un proceso de diseño con un sinfín de ligeros cambios y mejoras. Supongamos ahora que este proceso ya se había estabilizado, dando como resultado un reglamento ideal, que funcionaba de modo impecable; cualquier persona que siguiera el reglamento sería un excelente ascensorista (he conseguido encontrar en línea uno de estos antiguos reglamentos, una publicación del ejército estadounidense, cosa que no es sorprendente, teniendo en cuenta su papel pionero en la instrucción y la práctica; en la figura 4.1 reproducimos una de sus páginas).

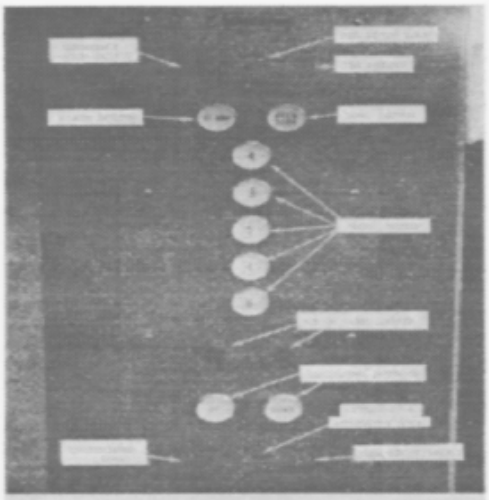


Figure 25. Elevator Control Panels.

numbers given. Be sure buttons for all stops requested are pressed before doors are closed.

(3) Say, "Next car, please," if more than maximum number of passengers attempt to enter car.

(4) Say, "Step back in car, please," in order to prevent crowding at car door.

(5) Ask passengers to, "Face front, please," if car is crowded and passengers are facing back or side of car.

4.2.2.2 Approaching Floor. As elevator approaches floor, operator should:

(1) Announce, "First floor," "Second floor," etc, as car slows to stop.

(2) Announce, "Street floor," as well as floor number, as, "First, street floor." This is necessary particularly in case of buildings on grade where street floor at one end is on different level from street level at other end of building.

4.2.2.3 As Car Stops. As car stops operator should:

(1) Say, "Please wait until car stops," if passengers attempt to alight from or enter while it is still leveling.

(2) Say, "Step up, please," or "Step down, please," if car does not stop level with landing sill. This is important as few people watch door sill when car stops.

4.2.3 Operating Procedures.

4.2.3.1 General:

(1) Parked elevator is never placed in service except under direction of supervisor.

(2) When at main floor, operator stands at attention well within the car.

(3) Operator never steps outside the car except when relieved from duty. Relieving operator steps into car and takes over control before dismissed operator leaves. Passengers are never allowed to remain in car without operator.

(4) When more than one car in bank is at main floor terminal, operators in cars other than next car to be loaded should close gates, and extinguish car lights.

(5) Cars should never be overloaded. Certificate of inspection is authority for weight load or number of persons permitted to ride in elevator.

(6) Floor signals are not passed without instructions from supervisor, unless car is full and signal "Transfer" switch is thrown.

(7) Passengers should not be hurried. It is both dangerous and discourteous.

(8) Operators never give information or make statements, either written or verbal, in connection with accidents occurring in the building. If statements are to be made, they must be given in presence of building manager or supervisor.

(9) When the car is out of service, the control mechanism is left inoperative by pulling "Emergency Switch." Where a motor generator is installed, supervisor shuts down set.

(10) Operators should make complete trips to top floor unless instructed otherwise

59

FIGURA 4.1. Página de un manual de instrucciones para ascensoristas

Ahora, imaginemos qué sucedió cuando un sencillo programa informático pudo encargarse de todas las tareas de control del ascensorista (en realidad, esto sucedió gradualmente, introduciendo diversos dispositivos mecánicos automáticos que se encargaban de las tareas menos especializadas del ascensorista, pero supongamos que este cambio de ascensoristas a sistemas completamente informatizados se realizó en un solo paso). El fabricante del ascensor reúne a un equipo de ingenieros de software (programadores) y les da el reglamento que han estado usando los ascensoristas: «Aquí tenéis las especificaciones, esto es lo que queremos; hacednos un programa que siga todas las normas de este reglamento tan bien como lo haría el mejor ascensorista y quedaremos satisfechos». Los programadores repasan el reglamento y

elaboran una lista de todas las acciones a realizar y de las condiciones bajo las cuales están permitidas o prohibidas; en el proceso, también pueden aprovechar para pulir algunos elementos ya innecesarios del reglamento; por ejemplo, si van a instalar sensores para garantizar que el ascensor se detiene siempre a la altura correcta, pueden eliminar el bucle que obliga al ascensorista a decir «Tengan cuidado con el escalón», o también pueden dejar en su lugar una grabación de voz que diga «Planta [n]; cuidado con el escalón».

El reglamento incluye normas acerca de cuántos pasajeros puede haber en el ascensor en cada momento, de modo que los programadores tendrán que enfrentarse a decisiones como «¿instalamos un torniquete para que el programa pueda ir contando cuántas personas entran y cuántas salen?» Seguramente no es una buena idea… quizá una báscula que determine el peso total de los ocupantes es un sistema mejor, más sencillo y menos intrusivo. En la ontología del ascensor, pues, en lugar de tener un sustantivo contable como pasajero u ocupante, habrá un sustantivo «de masa», como peso o carga. Metafóricamente, podemos decir que el ascensor se preguntará «cuánto peso» en lugar de «cuántas personas». De modo parecido, podemos señalar que un oso polar no intenta contar copos de nieve, sino que es consciente de la presencia o ausencia de nieve, o que el oso hormiguero lame un montón de hormigas con su lengua, a diferencia de un pájaro insectívoro, que sí que sigue la pista de cada insecto. Fijémonos que, del mismo modo que no tenemos que especular acerca de la conciencia del ascensor para marcar esta distinción, podemos considerar que los animales tienen diferentes ontologías sin tener que zanjar los problemas de si son conscientes de sus ontologías o simplemente se aprovechan de diseños que se pueden interpretar como poseedores de tales ontologías (interpretables así por ingenieros que aplican ingeniería inversa o ingeniería directa).

Volvamos a la ontología del ascensor. Para determinadas funciones le sirve el «peso», pero aún necesita controlar peticiones individuales a las que debe responder adecuadamente: «arriba» y «abajo» desde el exterior; «quinta planta» o «planta baja» y «abrir puertas» desde el interior. Y por cuestiones de seguridad, tiene que supervisarse a sí mismo, comprobar periódicamente el estado de sus diversos componentes y determinar si están funcionando bien y en las condiciones en que se supone que lo deben hacer. Ha de iluminar los botones cuando se pulsan y apagarlos cuando la tarea correspondiente al botón ya se ha completado (o por

otras razones, quizá). El grado de meticulosidad (o de obsesión-compulsión) del controlador puede variar, pero los programas que estén diseñados para ser descuidados con las interferencias o los fallos no se mantendrán mucho tiempo. Y si hay otros ascensores en un mismo vestíbulo (de un edificio de oficinas o un hotel), será relevante que los ascensores puedan comunicarse entre ellos o que haya un director que emita todas las órdenes (cuando se diseñan los ascensores de modo que usen referencias deícticas del tipo «¿dónde estás respecto a mi posición actual?», se mejora y se simplifica la cooperación entre los ascensores, sin necesidad de un controlador maestro omnisciente).

Es útil escribir el programa de control resultante en pseudocódigo, una especie de lengua criolla, mezcla del lenguaje humano cotidiano y el código fuente más exigente. Una línea de pseudocódigo podría ser algo así: SI PLANTALLAMADA > PLANTAACTUAL, ENTONCES SUBIR HASTA PLANTALLAMADA = PLANTAACTUAL Y PARAR; ABRIRPUERTAS; ESPERAR.

En cuanto el programa ya está claro en pseudocódigo y parece ser lo que se ha pedido, se puede traducir el pseudocódigo a código fuente, que es un sistema mucho más estructurado y riguroso de operaciones, con definiciones de términos (variables, subrutinas, etc.). El código fuente es relativamente legible por los humanos (al fin y al cabo, son los humanos quienes lo escriben) y, por tanto, las reglas y los términos del reglamento están representados en él de manera bastante explícita, si uno sabe dónde buscarlos. Hay dos rasgos que aún facilitan más esta tarea: en primer lugar, los nombres de las variables y operaciones se suelen escoger de modo que su significado sea claro (PLANTALLAMADA, PESOTOTAL, ABRIRPUERTAS, etc.); en segundo lugar, los programadores pueden añadir comentarios al código fuente, explicaciones que permiten a otros lectores humanos saber qué quería hacer el programador y qué hacen las diversas partes del código. Al programar, es una buena práctica dejar comentarios, ya que incluso el propio programador puede olvidar qué es lo que quería hacer con esa línea concreta de código; es más, al corregir errores de programación, estos comentarios resultan muy útiles. El código fuente debe estructurarse cuidadosamente, con una sintaxis muy estricta y con cada elemento en el lugar correcto, ya que luego es procesado por un compilador que traduce ese código fuente a un conjunto de secuencias de operaciones básicas ejecutables por la máquina real (o virtual). Un

compilador no puede suponer lo que un programador quería decir en una línea de código, de modo que el código debe poder decir al compilador exactamente qué operaciones hay que realizar; el compilador, por su parte, puede realizar estas tareas de formas muy diferentes y debe poder determinar la más eficaz en cada situación.

En el pseudocódigo, junto con miles de otras instrucciones, habrá una con este aspecto:

```
SI PESO-EN-KILOS > n ENTONCES PARAR; ABRIRPUERTAS


% Evita que el ascensor se mueva si se supera el peso máximo
autorizado


% Cuando alguien sale y el peso se reduce, continúa el funcionamiento
normal
```

Las frases tras los «%» son comentarios que desaparecen cuando se compila el código fuente. Los términos en mayúsculas también desaparecen del código que el compilador envía al chip informático que ejecuta el programa; solo sirven para que los programadores recuerden qué variable es cada una, y PESO-EN-KILOS está escrito así para recordar que el número que ponen en el programa para el peso máximo autorizado tiene que estar en kilogramos (en 1999, la sonda de la NASA Mars Climate Orbiter, de 125 millones de dólares, fracasó en su misión a Marte porque, para la distancia al planeta, una parte del sistema de control usaba metros y otra parte, pies; el resultado fue que la sonda se acercó demasiado y quedó destruida; sí, todos cometemos errores). En resumen, los comentarios y los nombres de las variables nos ayudan a entender el fundamento de diseño del sistema, pero son invisibles para el hardware y las máquinas. Cuando el programa está acabado y probado y se considera satisfactorio, la versión compilada puede grabarse en la ROM (memoria de solo lectura) para que la CPU (unidad central de

proceso) pueda acceder a ella. Las «reglas» que resultaban tan explícitas y evidentes en las primeras etapas del proceso de diseño han quedado implícitas en los ceros y unos que leerá el hardware.

El objetivo de toda esta divagación sobre los aspectos básicos de la programación es que el ascensor final tiene unos curiosos parecidos con los seres vivos y, a la vez, profundas diferencias. Para empezar, sus acciones se adecúan muy bien a las circunstancias; es un buen ascensor, que hace siempre las acciones correctas; casi podríamos decir que es inteligente (como los mejores ascensoristas de antaño). En segundo lugar, esta excelencia es debida a que su diseño contiene la ontología correcta; emplea variables que vigilan todas las características del mundo que es relevante para realizar la tarea e ignora todas las demás (si los pasajeros son jóvenes o viejos, muertos o vivos, ricos o pobres…). En tercer lugar, no necesita saber cuál es su ontología ni por qué es la que es. La motivación del programa es algo que solamente tienen que entender sus diseñadores; y tienen que hacerlo a causa de la naturaleza del proceso de I+D que ha construido el programa final: es un proceso de diseño inteligente (bastante inteligente). He aquí la profunda diferencia que debemos aclarar al pasar a la ontología de seres vivos simples, productos de la evolución por selección natural y no de un diseño inteligente.

Las bacterias son buenas en mantenerse con vida: hacen las acciones correctas y controlan las cosas que les son más relevantes; los árboles y los champiñones son igual de inteligentes o, si queremos decirlo con más precisión, están inteligentemente diseñados para hacer las acciones correctas en el momento adecuado. Todos ellos tienen mentes del tipo «ascensor», no mentes «elevadas» como nosotros;[5] tampoco las necesitan. Y sus mentes de ascensor son (tienen que ser) el producto de un proceso de I+D con pruebas y errores que, poco a poco, ha estructurado su maquinaria interna para que pase de un estado a otro de un modo que con gran probabilidad (aunque no al cien por cien) sirve a sus limitados pero vitales intereses. A diferencia del ascensor, su maquinaria no está diseñada por diseñadores inteligentes que han estudiado, discutido y pensado los porqués de sus componentes, de modo que no hay nada, nada en absoluto, que haga el papel de los comentarios y los nombres de variables de un programa. Este es el punto clave de la transformación que lograron Darwin y Turing con su extraña inversión del razonamiento.

Los ascensores pueden hacer cosas muy inteligentes: optimizan sus recorridos, con el consiguiente ahorro de tiempo y energía, ajustan automáticamente la velocidad para reducir la incomodidad de los pasajeros, «piensan en todo» lo que hay que pensar, obedecen las instrucciones e incluso responden a las preguntas más frecuentes. ¡Los buenos ascensores se ganan su sueldo! Y todo esto lo hacen sin neuronas ni órganos de los sentidos ni dopamina ni glutamato ni ningún otro componente orgánico de los cerebros. Así, parece justificado decir que lo que hacen tan inteligentemente es un ejemplo perfecto de competencia sin la más mínima pizca de comprensión ni de conciencia. A no ser, claro está, que consideremos que la maquinaria que le proporciona esta limitada capacidad tenga una pizca, o quizá incluso dos pizcas, de comprensión (siguiendo esta línea, su sistema de supervisión se puede juzgar como un paso muy básico hacia la conciencia).

Tanto si queremos otorgar un ápice de comprensión al ascensor como si no, estamos obligados a hacer lo mismo con las bacterias, con los árboles y con los champiñones; todos muestran una extraordinaria capacidad de mantenerse con vida en sus limitados nichos mediante la maquinaria bien diseñada que incorporan, gracias a sus genes. Ahora bien, esta maquinaria fue diseñada por el proceso de I+D de la selección natural, de modo que no hay absolutamente nada en la historia de esta I+D, en ningún lugar ni en ningún momento, que represente las motivaciones de las funciones de grandes sistemas o de sus componentes del mismo modo que los comentarios y los nombres de variables representan estas mismas funciones para los diseñadores humanos. Sin embargo, las motivaciones están ahí, para ser descubiertas por ingeniería inversa. Casi podéis dar por sentado que hay una razón por la cual las partes están conformadas como lo están, por la cual los comportamientos se organizan del modo en que lo hacen, y esta razón será una «justificación» del diseño (o justificará un diseño previo que ahora se ha convertido en vestigial o se ha transformado por evolución a una nueva función). La justificación será directa en términos de ingeniería: si eliminas el elemento o lo remodelas, el sistema no funcionará tan bien o no funcionará en absoluto. Las afirmaciones acerca de estas motivaciones flotantes deben ser, y pueden ser, comprobables, y en muchos casos se confirman más allá de cualquier atisbo de duda.

Volvamos a nuestro ascensor, ya automatizado por completo. ¡Tachán! Un ser humano de carne y hueso (no un homúnculo metafórico) acaba de

ser sustituido por una máquina. Y la máquina sigue las mismas reglas que el ascensorista humano. ¿Seguro? Vale, de acuerdo, no exactamente, solo más o menos. Estamos aquí ante un interesante punto medio entre un ser humano que memoriza las reglas que dictan su comportamiento (y, por lo tanto, las representa literalmente en su mente y las consulta) y un planeta, cuya órbita se describe con elegancia mediante ecuaciones que él «obedece». A menudo, nosotros mismos también nos situamos en este punto medio, cuando, a base de práctica, hemos internalizado y convertido en rutina una serie de reglas explícitas que luego podemos descartar o incluso olvidar (como «llevan tilde las palabras agudas acabadas en vocal, en n o en s»). Y también se pueden seguir reglas que aún no se han expresado de modo explícito: la gramática castellana, por ejemplo, tiene reglas no explícitas que siguen desafiando a los lingüistas, que continúan dando vueltas y más vueltas para intentar redactar una versión totalmente satisfactoria de un «manual de instrucciones» para el español, cuando cualquier hablante de la lengua de diez años, de algún modo, tiene instalada y depurada una versión bastante buena del código ejecutable para hablar y entender el idioma.

Antes de ponernos con las mentes de los animales, me gustaría plantear algunos otros ejemplos de diseños de artefactos que nos ayudarán a enmarcar el problema que la evolución ha resuelto al diseñar animales competentes.

LOS DISEÑADORES INTELIGENTES DE OAK RIDGE Y DE LA GOFAI

Después de setenta años, aún hay secretos de la segunda guerra mundial por revelar. Los gigantescos logros de Alan Turing en Bletchley Park al romper los códigos alemanes de la máquina Enigma ya son conocidos y celebrados como merecen, si bien algunos detalles aún se consideran demasiado delicados como para sacarlos a la luz. En cambio, es muy posible que solamente los estudiantes de historia de la física y la ingeniería nuclear estén familiarizados con el papel que tuvo el general Leslie Groves en la exitosa culminación del Proyecto Manhattan. Solo se necesitaron seis años desde que, en agosto de 1939, el presidente

Roosevelt recibió la carta de Albert Einstein y Leo Szilard informándole de la posibilidad de una bomba atómica hasta que se produjo el lanzamiento de la primera bomba, sobre Hiroshima, el 6 de agosto de 1945. Los primeros tres años estuvieron dedicados a investigación básica y a estudios de viabilidad, durante los cuales casi todos los implicados sabían exactamente qué era lo que estaban intentando hacer. En 1942, Leslie Groves fue nombrado director de lo que posteriormente se denominaría Proyecto Manhattan y, en tres años, bajo una enorme presión y combinando más I+D con la gigantesca (y novedosa) tarea de refinar uranio para uso militar, se contrataron, formaron y asignaron miles de trabajadores, la mayoría de los cuales destinados a controlar los nuevos aparatos inventados para separar el isótopo 235 del uranio, que representa tan solo un 1% del uranio 238 previamente refinado.

En el punto álgido de las operaciones, trabajaban en el proyecto más de 130.000 personas a jornada completa, pero solo una diminuta parte tenía claro qué es lo que estaban haciendo. ¡Eso sí que era competencia sin comprensión! El principio de mínimo conocimiento fue impuesto en grado sumo. En el K-25, la planta de difusión gaseosa, en la ciudad prefabricada de Oak Ridge, en Tennessee, decenas de miles de hombres y mujeres trabajaban sin descanso ocupándose de mandos, botones y palancas en una tarea que llegaron a efectuar con exquisita habilidad pero sin una pizca de comprensión. Sus reacciones tras Hiroshima lo dejaron claro: no sabían si estaban construyendo piezas para aviones, refinando aceite para cigüeñales de submarinos o cualquier otra cosa. Pensemos en la planificación que se necesita para crear un sistema de formación que convierte a los trabajadores en expertos sin que sepan en qué son expertos. El grado de secretismo fue superior al que se había establecido nunca antes (y quizá también, nunca después). Por supuesto, Leslie Groves y los planificadores tenían que saber todos los detalles del proyecto; eran diseñadores inteligentes, equipados con una comprensión exacta de los requisitos de la tarea a realizar, y solo mediante esta comprensión pudieron crear el entorno resguardado lleno de competencia sin comprensión.

El proyecto demostró una cuestión más allá de toda duda: se pueden crear niveles muy fiables de competencia prácticamente sin comprensión para tareas muy aisladas. Por lo que sé, incluso hoy en día, la distribución exacta de comprensión entre toda la mano de obra del Proyecto Manhattan sigue siendo un secreto bien guardado. ¿Qué

necesitaban saber los ingenieros y arquitectos que diseñaron el edificio K-25? En el momento en que lo construyeron, en unos pocos meses, era el edificio más grande del mundo. Algunos tenían que saber para qué se usarían los kilómetros y kilómetros de tuberías especializadas, sin duda, pero seguramente los que diseñaron el tejado, los cimientos o las puertas, no tenían la más mínima idea de lo que llevaban entre manos. No deja de ser interesante pensar en que, mientras Groves y su equipo diseñaban inteligentemente un sistema de miles de tareas humanas que requerían una comprensión mínima, al otro lado del Atlántico, Turing y sus compañeros estaban diseñando inteligentemente un sistema que podía sustituir esos ciegos homúnculos por componentes electrónicos. Pocos años después, muchos científicos e ingenieros, la mayoría de los cuales habían participado en uno u otro de estos dos innovadores proyectos bélicos, empezaron a aprovechar la creación de Turing de ladrillos competentes sin comprensión para fundar el atrevido campo de la inteligencia artificial.

Turing, en 1950, profetizó que al cambiar de siglo «el uso de las palabras y la opinión culta general habrá cambiado de tal modo que se podrá hablar de máquinas pensantes sin que nadie se muestre en desacuerdo». Los primeros trabajos en este campo fueron brillantes, coyunturales, ingenuamente optimistas y, por qué no decirlo, llenos de arrogancia. Sin duda, una inteligencia artificial tenía que poder ver, además de pensar, de modo que concibamos primero una máquina que vea; el tristemente célebre «proyecto de verano sobre la visión» que emprendió el MIT a mediados de la década de los sesenta fue un intento de «solucionar el problema de la visión» en unas vacaciones, dejando los problemas más difíciles para después. Para los patrones actuales, los «cerebros electrónicos gigantes» con los cuales se llevaron a cabo los primeros trabajos de inteligencia artificial, eran insignificantes y desesperadamente lentos, lo que tenía el efecto secundario de convertir la eficiencia en una prioridad; nadie se molestaría en elaborar un modelo informático que llevara días para responder a un conjunto realista de datos de entrada, especialmente si el objetivo era aprovechar la enorme velocidad de una computadora para tratar problemas reales al instante.

Los primeros trabajos en inteligencia artificial, o GOFAI (good old-

fashioned artificial intelligence, como la bautizó Haugeland [1985]), se basaban en un enfoque «de arriba abajo» e «intelectualista» de la

inteligencia artificial: anotar lo que los expertos humanos saben, en un lenguaje que la computadora pueda manipular con máquinas inferenciales que exploren los enormes bancos de memoria repletos de ese «conocimiento mundial» y, a partir de ahí, deducir los teoremas que se necesitarían para tomar decisiones fundadas y controlar adecuadamente las eventuales «extremidades» efectoras que esa inteligencia artificial pudiera tener. Retrospectivamente, la GOFAI se puede ver como un esfuerzo para construir un ente bastante cartesiano, una especie de experto racionalista con incontables proposiciones almacenadas en su memoria y con todo el conocimiento incluido en su capacidad para extraer conclusiones a partir de los axiomas relevantes y detectar contradicciones en su conocimiento del mundo, y todo de la manera más eficiente posible. Al fin y al cabo, ¿acaso un agente inteligente no es un ser racional con conocimiento de causa, que puede pensar con rapidez, empleando las proposiciones que conoce, para planificar acciones y así enfrentarse a las contingencias que puedan surgir? En ese momento, este enfoque parecía una buena idea, y para algunos investigadores en este campo, aún lo es.[6]

La insistencia en la velocidad y la eficiencia obligó a trabajar primero en «problemas de muestra». Buena parte de estos problemas a pequeña escala se solucionaron en mayor o menor grado, y sus soluciones han tenido aplicación en el mundo no especialmente exigente de sistemas de control en entornos restringidos (de ascensores y lavaplatos a refinerías y aviones), en diagnóstico médico, juegos y otros campos de investigación o interacción bien delimitados: reservas de vuelos, corrección ortográfica (e incluso gramatical), etc. Podemos considerar estas creaciones como descendientes más o menos indirectos de los sistemas aislados y compartimentados creados por Groves y su equipo de diseñadores inteligentes, sometidos al principio de mínimo conocimiento y basados en la capacidad de los diseñadores para idear sistemas formados por subsistemas equipados con las competencias precisas que necesitarán para resolver los problemas a los que se puedan enfrentar. Por muy brillantes que fueran, los primeros diseñadores de inteligencias artificiales no eran omniscientes (y el tiempo era lo principal), de modo que redujeron el abanico y la variedad de datos que cada subsistema tenía que aceptar y manipular, creando programas con miles de puestos de trabajo resguardados para proteger a los «sabios bobos» (las subrutinas) que trabajaban en ellos.

En todo este proceso se aprendió mucho y se inventaron y refinaron prácticas y técnicas que ayudaron a visualizar cuán difícil es la tarea de diseñar una mente humana independiente, imaginativa e ilimitada. El sueño de una enciclopedia andante, un sabelotodo de eficiencia burocrática, organizado «de arriba abajo» y programado a mano no ha desaparecido del todo, pero a medida que las dimensiones de un proyecto así han ido quedando más claras, se ha producido un saludable cambio hacia una estrategia diferente: emplear las colosales cantidades de macrodatos (el big data) y las nuevas técnicas estadísticas de búsqueda y reconocimiento de patrones en los campos de extracción de datos y aprendizaje profundo para obtener trabajosamente la información necesaria con un enfoque más «de abajo arriba».

Aún me quedan muchas cosas por decir sobre estos avances, pero de momento, la idea que tenemos que identificar es que el formidable aumento en la velocidad y la potencia de los ordenadores a lo largo de los años ha abierto las puertas a poder explorar procesos de extracción de información «ineficientes», «sin sentido», «ciegos», menos «burocráticos» y más parecidos a los procesos evolutivos, con unos resultados impresionantes. Gracias a estos nuevos enfoques, ahora podemos reflexionar detalladamente acerca de cómo pueden haber evolucionado los sistemas comparativamente simples que controlan las bacterias, los gusanos o las termitas, mediante los procesos de la selección natural, unos procesos «de abajo arriba», sin previsión y de fuerza bruta. Dicho de otro modo, queremos descubrir cómo la evolución puede hacer el papel de Leslie Groves para organizar operarios inconscientes en equipos eficaces, pero sin el beneficio de la comprensión y la previsión de que disponía Groves.

El diseño inteligente de arriba abajo funciona. La política de planificar por adelantado, articular los problemas, refinar las tareas y representar con claridad las razones de cada paso es una estrategia que no solamente ha parecido evidente a los inventores y solucionadores de problemas durante milenios, sino que ha quedado demostrada en incontables triunfos de la previsión y del ingenio en todos los campos del esfuerzo humano, desde la ciencia y la ingeniería hasta la política, la gastronomía, la agricultura y la navegación. Antes de Darwin se consideraba la única manera de lograr un diseño; el diseño sin un diseñador inteligente se juzgaba imposible. Pero resulta que este diseño de arriba abajo, en realidad, es el responsable de mucho menos diseño en el mundo de lo

que se suele creer, y para algunos de los «logros de la habilidad creadora», repitiendo una vez más las palabras de Beverley, este enfoque ha resultado inútil. La extraña inversión del razonamiento de Darwin y la inversión de Turing, igualmente revolucionaria, son aspectos de un mismo descubrimiento: la competencia sin comprensión. La comprensión, lejos de ser un talento divino a partir del cual debe surgir todo diseño, es un efecto emergente de los sistemas con competencia sin comprensión: la selección natural, por un lado, y la computación ciega, por otro. Estas dos ideas se han comprobado más allá de cualquier duda razonable, pero en algunos ambientes todavía generan consternación e incredulidad, sensaciones que he intentado desvanecer en este capítulo. Los creacionistas no encontrarán ningún código con comentarios en los mecanismos internos de los organismos ni los cartesianos hallarán ninguna res cogitans inmaterial «allí donde se produce toda comprensión».

** Se refiere a la Oda a un ruiseñor de John Keats, uno de los poemas más famosos de la literatura en lengua inglesa. (N. del t.)

5

LA EVOLUCIÓN DE LA COMPRENSIÓN

ANIMALES DISEÑADOS PARA OCUPARSE DE OFRECIMIENTOS

Está claro que los animales han sido diseñados mediante la selección natural, pero esta declaración de confianza en la evolución no resulta muy informativa. Más específicamente, ¿cómo puede la evolución conseguir esto? Uno de los resultados de nuestro paréntesis acerca de cómo diseñar un controlador de ascensores y toda su parentela de artefactos es una apreciación más aguda de la diferencia entre los procesos de I+D y la evolución por selección natural. La computadora en la que los diseñadores (programadores) prueban y ejecutan sus soluciones es ella misma un producto del diseño inteligente, tal como hemos señalado, y su conjunto inicial de ladrillos competenciales (aritmética y ramificación condicional) invita a los potenciales programadores a idear también sus tareas en un esquema de arriba abajo, como un proceso de resolución de problemas en que intentan incrustar su conocimiento del problema en las soluciones que conciben.

No sería extraño preguntar «pero ¿cómo podría ser de otro modo?» El diseño inteligente empieza con un objetivo (que puede refinarse o incluso abandonarse por el camino) y trabaja de arriba abajo: los diseñadores utilizan todo lo que conocen para guiar su búsqueda de soluciones a los problemas de diseño (y subproblemas, subsubproblemas, etc.) que se plantean. En cambio, la evolución no tiene objetivos, no tiene problemas predefinidos, no tiene conocimientos a aportar en esta tarea; no, la evolución va dando tumbos, miope y sin dirección, con lo que ya ha creado, probando ciegamente ajustes y variaciones y conservando aquellos que se demuestran útiles o, como mínimo, no muy perjudiciales.

¿Puede algo tan sofisticado intelectualmente como una computadora digital haber evolucionado por un proceso de selección natural de abajo arriba? Algo así es muy difícil de imaginar, o incluso de considerar seriamente, lo que ha llevado a algunos pensadores a concluir que, dado que la evolución no puede haber creado una computadora (ni un programa informático para ejecutarse en ella), las mentes humanas tampoco pueden ser producto exclusivo de la selección natural y las aspiraciones de la inteligencia artificial no son más que una quimera. El físico y matemático Roger Penrose (1985) es el representante más ilustre de esta postura. En pro del argumento, supongamos que, efectivamente, la evolución por selección natural no pudo hacer evolucionar directamente una computadora digital viviente (un árbol-máquina de Turing o una tortuga-máquina de Turing, por ejemplo); aun así, hay una vía indirecta: dejemos que la selección natural haga evolucionar en primer lugar mentes humanas y luego, que estas mentes humanas diseñen inteligentemente Hamlet, la Sagrada Familia y la computadora, entre otras incontables maravillas. A primera vista, este proceso de arranque parece casi mágico e incluso contradictorio: ¿acaso no son Shakespeare, Gaudí o Turing «creaciones» mucho más brillantes y esplendorosas que cualquiera de sus productos? En cierto modo sí, sin duda, pero también es cierto que sus productos tienen características que no podrían haber aparecido sin sus artífices.

Si aterrizarais en un planeta lejano y estuvierais buscando señales de vida por una orilla, ¿qué os entusiasmaría más, una almeja o un rastrillo para almejas? Una almeja tiene millones de complejas partes móviles, mientras que el rastrillo tiene apenas un par de toscas piezas fijas; pero el rastrillo tiene que ser el artefacto de algún ser vivo, de algo mucho más impresionante que una almeja. ¿Cómo puede ser que un proceso lento y ciego consiga construir algo que puede construir algo que un proceso lento y ciego no podría construir por sí mismo? Si os parece que esta pregunta no tiene respuesta, que es una pregunta retórica, es que aún estáis encantados por el hechizo que Darwin logró romper, que aún no estáis en posición de asumir su extraña inversión del razonamiento. Y ahora, precisamente, podemos ver cuán extraña y radical es esta inversión: se trata de un proceso sin diseñador inteligente que puede crear diseñadores inteligentes que pueden diseñar cosas que nos permiten comprender cómo un proceso sin diseñador inteligente puede crear diseñadores inteligentes que pueden diseñar cosas.

Los pasos intermedios son instructivos. ¿Qué es lo que nos revela que el rastrillo es artificial? Su simplicidad, que indica su dependencia de alguna otra cosa para poder desafiar la segunda ley de la termodinámica, manteniéndose como un conjunto de átomos uniforme y simétrico, yuxtapuestos de modos improbables. Algo tuvo que reunir y refinar este conjunto; algo complejo.

FIGURA 5.1. Rastrillo para almejas

Volvamos una vez más a los organismos simples. La idea de que cualquier organismo tiene su ontología (en el mismo sentido que el ascensor) ya fue prefigurada por Jakob von Uexküll (1934) con su concepto del Umwelt de un organismo, el entorno conductual formado

por todas las cosas relevantes para su bienestar. Una idea parecida es el concepto de ofrecimientos del psicólogo J. J. Gibson (1979): «lo que el entorno ofrece a un animal, para bien o para mal». Los ofrecimientos son las oportunidades relevantes que hay en el entorno de cualquier organismo: cosas para comer o con las que aparearse, orificios por los que pasar o por los que mirar, agujeros en los que esconderse, objetos a los que encaramarse, etc. Ni Uexküll ni Gibson dijeron nada acerca de si la conciencia (como algo aún por definir) estaba implicada en el hecho de tener un Umwelt poblado por ofrecimientos, pero como los casos estudiados por Uexküll incluían amebas, medusas y garrapatas, parece que, al igual que Gibson, estaba más interesado en caracterizar los problemas a los que los organismos se enfrentan y solucionan que en determinar cómo se llevan a cabo internamente tales soluciones. El Sol forma parte de la ontología de una abeja: su sistema nervioso está diseñado para aprovechar la posición del Sol en sus actividades; las amebas y los girasoles también tienen al Sol en sus respectivos Umwelten, pero a falta de sistema nervioso, emplean otros mecanismos para responder adecuadamente a su posición. Igualmente, el concepto que tenga el ingeniero de la ontología del ascensor es todo lo que necesitamos para empezar y podemos dejar para más tarde las cuestiones acerca de si, cuándo y por qué la ontología de un organismo, o de una estirpe de organismos, pasa de algún modo a ser manifiesta en su conciencia y no simplemente implícita en las respuestas diseñadas de su maquinaria interna. Es decir, los organismos pueden ser beneficiarios de rasgos de diseño que implican la existencia de una ontología pero sin que haya ninguna necesidad de que tales organismos representen dichas ontologías en ningún sentido más fuerte (de modo consciente, semiconsciente o inconsciente). La forma del pico de un pájaro, junto con algunos otros rasgos anatómicos secundarios, implica una dieta de semillas duras, de insectos o de peces de modo que podemos abastecer los Umwelten de las diversas especies de aves con semillas duras, insectos y peces, como ofrecimientos específicos de la especie, solamente a partir de estos rasgos anatómicos, aunque, por supuesto, no está de más confirmar tal implicación estudiando su comportamiento, siempre que se pueda. Pero la forma del pico no es una representación (en ningún modo relevante) de su alimento preferido ni de su modo de obtenerlo.

Los paleontólogos obtienen conclusiones acerca de las preferencias predatorias y otros comportamientos de especies extinguidas mediante esta forma de inferencia y pocas veces se señala que esto depende de asumir suposiciones adaptacionistas acerca de los diseños de las criaturas

fósiles. Planteemos el ejemplo propuesto por Niles Eldredge (1983) sobre los trabajos de Fisher (1975) acerca de la velocidad de nado de los cangrejos cacerola; Eldredge los cita para mostrar que plantear la pregunta histórica «¿qué ha pasado?» (el cómo es que) es una estrategia mejor que plantear la pregunta adaptacionista (el para qué), con sus suposiciones de optimización. Pero la conclusión de Fisher sobre la rapidez con que nadaban los extintos cangrejos cacerola

depende de una suposición adaptacionista muy prudente acerca de lo que es bueno: cuanto más rápido, mejor (dentro de unos límites). La conclusión de que los cangrejos cacerola del Jurásico nadaban más rápido depende de la suposición de que, teniendo en cuenta su forma, lograrían la velocidad máxima nadando con un ángulo determinado, y de que nadarían así para lograr la máxima velocidad. De este modo, [Fisher necesita] hacer uso de consideraciones de optimización en absoluto controvertidas, incluso tácitas, para tener una idea de «lo que pasó» hace 150 millones de años. (Dennett 1983)

Recordad, la biología es ingeniería inversa, y la metodología de la ingeniería inversa está comprometida con consideraciones de optimización. La pregunta «¿para qué es (o era) buena esta característica?» siempre está en la punta de la lengua; sin ella, la ingeniería inversa se diluye en mera confusión.

Tal como he dicho en el primer párrafo del libro, las bacterias no saben que son bacterias, pero sin duda reaccionan ante otras bacterias al modo bacteriano de hacerlo, y pueden evitar, rastrear o vigilar elementos que ellas distinguen en su Umwelt, sin ninguna necesidad de saber lo que están haciendo. Las bacterias pertenecen a la ontología de las bacterias del mismo modo que los pisos y las puertas pertenecen a la de los ascensores; la única diferencia es que las bacterias son mucho más complejas. Igual que hay razones por las que los circuitos de control de un ascensor están diseñados del modo en que lo están, también hay razones por las que las redes de control de proteínas de una bacteria están diseñadas del modo en que lo están: en ambos casos, los diseños han sido optimizados para enfrentarse a los problemas de un modo eficiente y eficaz.[1] La diferencia principal es que el diseño de los circuitos del

ascensor es obra de diseñadores inteligentes que han elaborado descripciones de los problemas y representaciones de soluciones razonadas, con todas las justificaciones. En el historial de I+D de las bacterias no había ningún código fuente, nunca se introdujeron comentarios para dar pistas de lo que pretendía la madre naturaleza. Pero ello no es óbice para que los biólogos evolutivos asignen funciones a determinadas características resultado de la evolución (las patas palmeadas son para propulsarse en el agua) e interpreten otras como errores de la naturaleza (un ternero de dos cabezas). De forma parecida, los editores de textos literarios de autores fallecidos hace tiempo no tienen que confiar en revelaciones autobiográficas dispersas por los manuscritos del autor para interpretar algunos fragmentos dudosos como deliberadamente engañosos y otros como simples errores tipográficos o lapsus de memoria.

El desarrollo de software es un campo relativamente nuevo del conocimiento humano. Aunque aún se halla en su infancia, se han identificado y corregido muchos problemas y puntos débiles y ha surgido toda una babel de lenguajes de programación, junto con una miríada de herramientas de programación para facilitar el trabajo. Aun así, la programación continúa siendo un «arte» e incluso en los programas comerciales publicados por las mejores empresas siempre acaban apareciendo errores que obligan a publicar actualizaciones para corregirlos. ¿Por qué no se ha automatizado la tarea de depuración de fallos para, así, eliminar cualquier posible error de buen principio? Los diseñadores humanos más inteligentes, plenamente conscientes de los propósitos del software, consideran la depuración del código una tarea abrumadora, incluso cuando pueden trabajar sobre un código con comentarios detallados y elaborado de acuerdo con prácticas recomendadas estrictamente reglamentadas (Smith 1985, 2014). Hay una razón por la que la depuración no se puede automatizar totalmente: lo que cuenta como fallo depende de todos los objetivos (y subobjetivos y subsubobjetivos) del software, y, al fin y al cabo, especificar con el detalle suficiente cuáles son estos objetivos (para proporcionarlos al eventual programa de depuración automática) cuesta el mismo trabajo que escribir el código ya depurado de buen principio, al menos a efectos prácticos.[2] Escribir y depurar código informático para proyectos ambiciosos es una de las pruebas más exigentes de la perspicacia humana, y en cuanto algún brillante programador concibe un nuevo recurso que libera al codificador de parte del arduo trabajo, subimos el listón de lo que esperamos que un programador pueda crear y probar. No es que esto sea un fenómeno sin

precedentes en la historia de las actividades humanas; la música, la poesía y las demás artes siempre han mostrado al potencial creador espacios ilimitados de posibles «acciones» que no se reducen por el hecho de inventar la notación musical, la escritura o las pinturas preparadas, ni la creación artística se vuelve rutinaria al incorporar sintetizadores y archivos MIDI, procesadores de texto y correctores ortográficos o imágenes digitales de alta resolución y millones de colores.

Y la naturaleza, ¿cómo depura sus diseños? Como no hay ningún código fuente ni comentarios para leer, no se puede producir ninguna depuración con explicaciones intelectualmente brillantes; en la naturaleza, la revisión de los diseños debe seguir un método mucho más derrochador: publicar y probar muchas variantes y dejar que las peores mueran, sin examinarlas. Este proceso no tiene por qué hallar el diseño globalmente óptimo, pero las mejores versiones accesibles localmente medrarán y las pruebas subsiguientes cribarán aún más a las mejores, de modo que para la próxima generación el listón estará un poco más arriba.[3] Tal como subraya el magnífico título de Richard Dawkins (1986), la evolución es un relojero ciego y, teniendo en cuenta su método de I+D, no es sorprendente que en sus productos abunden recovecos oportunistas y miopes pero retorcidamente eficaces (¡excepto cuando no lo son!). Uno de los distintivos del diseño por selección natural es que está lleno de errores, en el sentido de un programador informático (bugs): fallos que solo se revelan en situaciones muy improbables, situaciones que nunca se dan en los procesos finitos de I+D que completan el diseño y, por tanto, tras generaciones de retoques aún no se han parcheado ni esquivado. Los biólogos son muy buenos sometiendo los sistemas que estudian a condiciones muy improbables, imponiéndoles retos extremos para descubrir cuándo, dónde y cómo fallan, y por qué.

Al aplicar ingeniería inversa a un organismo, lo que los biólogos suelen descubrir es algo así como el «código espagueti» enmarañado e indescifrable de los programadores poco cuidadosos. Con algo de esfuerzo, al descifrar un código espagueti, en general podemos señalar qué posibilidades poco probables nunca se les ocurrieron a los diseñadores en su miope búsqueda de la mejor solución a los problemas planteados. ¿En qué estaban pensando? Cuando preguntamos la misma cuestión sobre la madre naturaleza, la respuesta es siempre la misma: en nada. No había ningún tipo de pensamiento, pero así y todo se las apañó para improvisar un diseño tan efectivo que ha sobrevivido hasta el día de

hoy, venciendo a la competencia en un mundo exigente, hasta que aparece algún biólogo sabiondo y nos revela sus puntos débiles.

Echemos un vistazo a los estímulos supranormales, un error de diseño que se puede encontrar en muchos organismos. Los experimentos de Niko Tinbergen (1948, 1951, 1953, 1959) con gaviotas mostraron un curioso sesgo en su maquinaria perceptual y conductual. La hembra adulta tiene una mancha roja en el pico que los polluelos pican instintivamente para estimular la regurgitación y, así, que la madre los alimente. ¿Qué pasaría si la mancha roja fuera más grande o más pequeña, más brillante o más tenue? Tinbergen mostró que los polluelos picaban con más facilidad a modelos ampliados de cartón de la mancha roja y, así, que estímulos supranormales provocan comportamientos supranormales. Tinbergen también mostró que las aves que ponen huevos de color azul claro y moteados de gris prefieren incubar huevos falsos de color azul brillante y con lunares negros, tan grandes que caen de ellos una y otra vez.

Una réplica habitual de los programadores es «¡no es un fallo, es una característica!», lo cual podría aplicarse también a los estímulos supranormales. Mientras su Umwelt no contenga a taimados biólogos con una imaginación desbordante que estimulan a las aves con artilugios artificiales, el sistema funciona muy bien, centrando el comportamiento del organismo en lo que (casi siempre) importa. La motivación flotante de todo el sistema es claramente lo bastante buena para todos los objetivos prácticos, y la madre naturaleza ha sido lo bastante sabia como para no despilfarrar en algo más seguro que, eventualmente, pudiera detectar el engaño. Esta filosofía de diseño se halla por todas partes en la naturaleza y ofrece las oportunidades en las que surgen las carreras armamentísticas, en que una especie se aprovecha de un atajo en el diseño de otra especie, lo que provoca un movimiento en el espacio de diseño que impulsa a ambas especies a desarrollar sistemas defensivos y ofensivos cada vez mejores. Las luciérnagas hembra se posan en el suelo observando las pautas de luz emitidas por los machos; cuando una hembra elige y devuelve la pauta de luz, el macho se abalanza para el apareamiento. Pero este ingenioso sistema de citas rápidas ha sido invadido por otra especie de luciérnaga, Photuris, que simula ser una hembra y atrae a los machos hacia su muerte. Photuris prefiere a los machos con señales más largas e intensas, de modo que los machos van evolucionando hacia cartas de amor más breves (Lewis y Cratsley 2008).

ANIMALES SUPERIORES Y SISTEMAS INTENCIONALES:

EL SURGIMIENTO DE LA COMPRENSIÓN

La competencia sin comprensión es el modo de trabajar de la naturaleza, tanto para sus métodos de I+D como para sus productos más simples y pequeños, diseños brillantes como las proteínas motrices, las enzimas de revisión, los anticuerpos y las células a las que dan vida. ¿Y qué hay de los organismos pluricelulares? ¿Cuándo surge la comprensión? Las plantas, desde las algas más diminutas a las enormes secuoyas, muestran muchas competencias aparentemente inteligentes: engañan a insectos, aves y otros animales para que les ayuden en su reproducción, forman provechosas alianzas con simbiontes, detectan valiosas fuentes de agua, siguen el Sol y se protegen de diversos depredadores (herbívoros y parásitos). Algunos especialistas sostienen (p. ej. Kobayashi y Yamamura 2003; Halitschke et al. 2008) que algunas especies de plantas, cuando son atacadas, pueden avisar a los individuos cercanos de una depredación inminente lanzando señales de aviso transportadas con el viento, con lo que permiten que los que reciben las señales intensifiquen sus mecanismos de defensa anticipadamente (aumentar la toxicidad o generar olores que ahuyentan a los depredadores o que atraen a simbiontes que los ahuyentan). Estas respuestas se dan con tal lentitud que son difíciles de observar como comportamientos propiamente dichos sin la ventaja que nos aporta la fotografía a cámara rápida, pero, al igual que el comportamiento microscópico de células aisladas, tienen motivaciones claras que los propios actores no tienen ninguna necesidad de entender.

Vemos aquí el surgimiento de algo parecido a un doble rasero de atribución. Es casi imposible describir y explicar estos procesos temporalmente organizados sin denominarlos comportamientos y explicarlos del mismo modo que explicamos nuestros propios comportamientos: señalando razones y suponiendo que están guiados por algo comparable a la supervisión perceptual, la recepción de información que activa, modula y da por finalizadas las respuestas. Cuando lo hacemos así, parece que atribuimos no solo competencia, sino también la comprensión que (en nosotros) suele acompañar a esta competencia conductual. Para comprender las plantas y las bacterias, las estamos

antropomorfizando. No es ningún pecado intelectual; es correcto denominar comportamientos a sus acciones, atribuir a los organismos tales competencias y explicar su existencia señalando motivaciones que tienen en cuenta las ventajas que, gracias a estas competencias, los organismos obtienen en su lucha por la supervivencia. Lo que estoy diciendo es que es correcto adoptar lo que he llamado el enfoque intencional. El único desliz es atribuir comprensión al organismo o a sus partes; afortunadamente, en el caso de las plantas y de los microbios, el sentido común interviene para impedir tal atribución: es bastante fácil entender cómo la maquinaria interna puede proporcionar su competencia sin que intervenga ningún tipo de mentalidad.

Podemos decir que los organismos que tienen competencias espectaculares sin necesidad de comprensión de sus propias motivaciones están «bien dotados». Son los beneficiarios de los talentos que les han sido concedidos, talentos que no son el producto de su esfuerzo o investigación particular; podemos decir, incluso, que han sido «bendecidos» con estos dones, no por Dios, por supuesto, sino por la evolución por selección natural. Si necesitamos un soporte para nuestra imaginación, podemos sostenernos en el estereotipo ya obsoleto del robot como un mecanismo ciego: las plantas no tienen comprensión, son robots vivientes (aquí aventuro una predicción: dentro de cien años esto se considerará un curioso resto del biocentrismo, un prejuicio trasnochado contra los robots con comprensión que había sobrevivido hasta bien entrado el siglo XXI).

Ya que estamos en el tema, es interesante recordar que, en el siglo XX, una de las objeciones más populares contra la GOFAI era:

La llamada inteligencia de estos programas no es más que la inteligencia (la comprensión) de los programadores. ¡Los programas no comprenden nada!

Voy a adoptar y adaptar esta idea, pero sin conceder (aún) comprensión a nadie ni a nada:

La llamada inteligencia de los árboles, las esponjas y los insectos no es suya; solo han sido diseñados con brillantez para hacer las acciones más adecuadas en el momento preciso y, aunque el diseño es brillante, el diseñador tiene la misma falta de comprensión que ellos.

Los adversarios de la GOFAI creían que no hacían nada más que afirmar lo que era obvio al plantear su crítica de las máquinas «inteligentes», pero fijaros en cómo se invierte el tirón emocional cuando hacemos la misma observación acerca de los animales. Si bien supongo que la mayoría de lectores se sentirán bastante cómodos con mi observación de que las plantas y los microbios simplemente están «bien dotados», bendecidos con competencias bien diseñadas pero, por lo demás, sin la más mínima comprensión, cuando aventuro esta misma opinión acerca de los animales «superiores», se me tachará de cruel y de aguafiestas.

Al fijarnos en los animales (especialmente los «superiores», como las aves o los mamíferos), la tentación de atribuir comprensión para describir y explicar las competencias es mucho más fuerte, y muchos insistirán en que es totalmente adecuada: ¡los animales sí que comprenden lo que están haciendo, mira cuán inteligentes son! Bueno, ahora que tenemos bien asido el concepto de competencia sin comprensión, debemos reconsiderar este benévolo punto de vista. Actualmente, se estima que más de la mitad del peso total de la vida del planeta (la biomasa) corresponde a bacterias y otros «robots» unicelulares, mientras que las plantas «robóticas» dan cuenta de más de la mitad del resto; luego tenemos a los insectos, con todas las termitas y hormigas ignorantes, que pesan más que toda la población humana celebrada por MacCready. Nosotros y nuestros animales domésticos formamos el 98% de la biomasa terrestre de vertebrados, pero esto es una parte muy pequeña de la vida del planeta. La competencia sin comprensión es el modo de vida de una inmensa mayoría de seres vivos del planeta, y debería ser la suposición de partida hasta que podamos demostrar que algunos organismos, en realidad, sí que comprenden lo que están haciendo, en un sentido u otro. Así, ahora la cuestión pasa a ser cuándo y por qué el diseño de organismos empieza a representar (o a incorporar con inteligencia de algún otro modo) las motivaciones flotantes de su maquinaria de supervivencia. Tenemos que corregir nuestras imaginaciones al respecto, ya que lo habitual es suponer que, siempre que en los animales «superiores» hay una motivación, habrá también algún tipo de comprensión.

Podemos presentar un ejemplo muy llamativo. Elizabeth Marshall Thomas es una experta y perspicaz observadora de animales (incluidos los animales humanos) y en uno de sus libros, The Hidden Life of Dogs (1993), se permite imaginar que los perros disfrutan de una sabia comprensión de sus actos: «Por razones que los perros conocen, pero nosotros no, muchas perras no se aparearán con sus hijos» (p. 76). Nadie cuestiona su resistencia instintiva a esta endogamia: seguramente se basan en los olores como pista, pero quién sabe qué otros elementos también contribuyen a dicha resistencia (he aquí un tema para investigaciones futuras). Pero la sugerencia de que los perros puedan tener un entendimiento de las razones que hay detrás de sus inclinaciones y comportamientos instintivos superior al que tenemos nosotros de las nuestras no es más que romanticismo desenfrenado. Estoy seguro de que Marshall Thomas es más sensata que esto; lo que quiero señalar es que cometió este desliz de forma natural, como una ampliación de la suposición generalizada, no como una hipótesis personal aventurada sobre el conocimiento que los perros tienen de sí mismos. Es como un antropólogo marciano que escribiera «por razones que los humanos conocen, pero nosotros no, muchos humanos bostezan cuando tienen sueño y alzan las cejas cuando se encuentran con un conocido». Hay razones para estos comportamientos, razones para qué, pero no son nuestras razones; podemos simular un bostezo o arquear las cejas por una razón concreta (dar una señal preestablecida, simular familiaridad con una persona atractiva pero desconocida, etc.), pero en condiciones normales, ni tan solo nos damos cuenta de que lo hacemos y, en consecuencia, no tenemos oportunidad de saber por qué lo hacemos (y, de hecho, aún no sabemos por qué bostezamos, aunque los perros no están en absoluto por delante de nosotros en esta investigación por mucho que bostecen tanto como nosotros).

¿Qué podemos decir de los comportamientos más claramente deliberados de los animales? Los cucos son parásitos de puesta, que no construyen sus propios nidos; en lugar de esto, la hembra del cuco pone su huevo a escondidas en el nido de una pareja anfitriona de alguna otra especie de ave, donde el huevo recibe las atenciones de sus involuntarios padres adoptivos. A menudo, la hembra del cuco hace rodar uno de los huevos de los anfitriones hasta que sale del nido, por si los padres son capaces de contarlos. En cuanto el polluelo cuco rompe el huevo (suele incubarse con más rapidez que los huevos de los anfitriones), dedica todos sus esfuerzos a lanzar fuera del nido los huevos restantes. ¿Por qué? Para maximizar los esfuerzos de crianza que recibirá de sus padres adoptivos. Los vídeos de

este comportamiento son una muestra escalofriante de una matanza eficaz y competente, pero no hay razón para suponer que exista lo que en términos legales llamaríamos dolo (una voluntad deliberada de causar daño, la mens rea del derecho consuetudinario anglosajón); el polluelo no tiene la más mínima idea de lo que está haciendo, aunque sin duda se beneficia de su propio comportamiento. Quizá podríamos echar un vistazo a una situación un poco más alejada del cleptoparasitismo de los cucos: ver cómo un pájaro construye el nido es una experiencia fascinante, en la que sin ninguna duda intervienen acciones muy especializadas de tejido e incluso costura (Hansell 2000). Se trata de un proceso con control de calidad y una pizca de aprendizaje. Los pájaros que nacen en cautividad, que nunca han visto cómo se construye un nido, cuando es el momento de construirlo, lo harán a partir de los materiales disponibles, creando un nido utilizable típico de su especie; el comportamiento, pues, es instintivo; ahora bien, en la siguiente estación construirán un nido mejor.

¿Cuánta comprensión atesora un pájaro constructor de nidos? Es un tema que los investigadores están estudiando (Hansell 2000, 2005, 2007; Walsh et al. 2011; Bailey et al. 2015), cambiando los materiales disponibles e interfiriendo de diversas maneras en las condiciones de construcción, para ver cuán versátiles e incluso previsores pueden ser los pájaros en cuestión. Teniendo en cuenta que la evolución solo puede aportar soluciones a los retos que surgen durante la I+D, podemos predecir que cuanto más novedosas sean las interferencias artificiales en el Umwelt del pájaro, menos probable será que este las interprete adecuadamente, a no ser que la estirpe de ese pájaro haya evolucionado en un entorno selectivo muy diverso que haya obligado a la selección natural a decidirse por diseños que, en lugar de estar totalmente programados, incorporan una gran plasticidad y mecanismos de aprendizaje. Es interesante señalar que cuando no hay bastante estabilidad en el entorno selectivo a lo largo del tiempo como para permitir que la selección natural «prediga» el futuro con precisión (que «seleccione» los mejores diseños para la siguiente generación), a esta le va mejor si dejan sin fijar, en parte, el diseño de la siguiente generación, como una computadora que pueda configurarse de muchas formas diferentes, en función de los gustos y hábitos del usuario.[4] El aprendizaje puede tomar las riendas allí donde la selección natural lo ha dejado, optimizando a los individuos durante su vida al obtener información del mundo con el que se encuentran y usar esta información para lograr mejoras locales. Pronto nos centraremos otra vez en esta vía

hacia la comprensión, pero antes me gustaría plantear unos pocos ejemplos más de comportamientos con motivaciones flotantes y sus implicaciones.

Quizá hayáis visto alguna vez vídeos de antílopes perseguidos por depredadores en extensas llanuras, y quizá os hayáis dado cuenta de que algunos hacen unos grandes saltos verticales durante su intento de huida. Esto se denomina salto de rebote. ¿Por qué lo hacen? Sin duda es un comportamiento ventajoso, porque los antílopes que practican el salto de rebote bien pocas veces resultan cazados. Se trata de una correlación causal que se ha observado detenidamente y que exige una explicación para qué. Ninguna descripción de todas las proteínas y otras moléculas de las células de los antílopes y los depredadores podrá revelar por qué se da esta correlación. Para llegar a una respuesta debemos echar mano del campo de la biología evolutiva denominado teoría de la señalización costosa (Zahavi 1975; Fitzgibbon y Fanshawe 1988). Los antílopes más fuertes y rápidos practican el salto de rebote para proclamar su buen estado físico a los perseguidores, como diciendo: «No os molestéis en cazarme, porque lo vais a tener crudo; centraos en alguno de mis parientes que no pueden saltar tanto como yo, ¡son una presa mucho más fácil!»; el cazador acepta esto como una señal honesta, difícil de falsear y, así, ignora al antílope saltarín. Es un acto de comunicación y un acto con una única motivación flotante, que ni el antílope ni el león tienen necesidad alguna de apreciar. Es decir, el antílope puede ser totalmente ignorante de por qué es buena idea practicar el salto de rebote si puede hacerlo, mientras que el león no tiene por qué comprender qué es lo que hace que considere poco atractivas las presas que saltan; ahora bien, si la señalización no fuera honesta y costosa, no podría mantenerse en la carrera armamentística evolutiva entre presa y depredador (si la evolución intentara una señal «barata», como sacudir la cola, que cualquier antílope puede emitir, sea fuerte o débil, a los leones no les saldría a cuenta fijarse en ella y, por tanto, no lo harían). Todo esto puede parecer una degradación aguafiestas y demasiado escéptica de la inteligencia del león y del antílope, pero es una aplicación estricta del mismo principio de ingeniería inversa que puede explicar el comportamiento del cuco, las termitas y las bacterias. En consecuencia, la regla de atribución tiene que ser que, si la competencia observada se puede explicar sin recurrir a la comprensión, no debemos permitirnos un antropomorfismo desmesurado. Si queremos atribuir comprensión debemos respaldarlo con demostraciones de un comportamiento mucho más inteligente. Como el salto de rebote no parece ser un elemento de un sistema más intrincado de

comunicación intraespecies o interespecies sobre diversos temas, las posibilidades de hallar una necesidad para algo que parezca comprensión son mínimas. Si creéis que este veredicto es demasiado escéptico, intentad imaginar experimentos que puedan demostrar que estáis en lo cierto.

¿Cómo podría un experimento respaldar el veredicto de la comprensión? Mostrando que los animales pueden hacer lo que nosotros, que tenemos comprensión, podemos hacer con variaciones de comportamiento. El salto de rebote es un tipo de alardeo o fanfarroneo, un comportamiento que nosotros también podemos ejecutar, pero nosotros también podemos fingir o contenernos de fanfarronear si se dan las condiciones en las que un comportamiento así pudiera ser contraproducente o aún peor. Podemos modular nuestro fanfarroneo, ajustándolo a diferentes receptores, o incluso podemos fanfarronear de un modo muy evidente y exagerado para comunicar que en realidad no estamos fanfarroneando, sino solo bromeando. Y así, indefinidamente. ¿Puede al antílope hacer algo parecido? ¿Puede contener sus brincos en situaciones en las cuales, por alguna causa novedosa, este comportamiento resultaría inadecuado? Si así fuera, sería una indicación de que tiene, y emplea, una cierta comprensión de las motivaciones de sus acciones.

Una motivación flotante bastante diferente rige el comportamiento de aves que nidifican en el suelo y fingen heridas, como el chorlitejo silbador; así, alejan a los depredadores del nido y los atraen hacia ellos mismos, fingiendo tener un ala rota y manteniéndose justo fuera del alcance del cazador hasta que están lo bastante alejados del nido. Esta técnica de distracción se puede observar en muchas especies de aves que nidifican en el suelo, muy distantes entre sí (Simmons 1952; Skutch 1976). Este comportamiento se suele considerar un engaño por parte del pájaro, y ciertamente lo parece: su objetivo es engañar al depredador. Podemos adoptar la útil técnica expositiva de Dawkins (1976) de inventar monólogos y crear uno para el chorlitejo:

Soy un pájaro que construye nidos en el suelo y no puedo proteger a mis polluelos de los depredadores que los descubran. Es muy posible que este depredador que se está acercando los descubra pronto si no lo distraigo; quizá su deseo de cazarme y comerme pueda distraerle, pero solo si cree que hay una razonable posibilidad de cazarme (no es tonto); y esto lo

creerá si le doy pruebas de que ya no puedo volar, lo que puedo hacer simulando que tengo un ala rota…

¡No está nada mal! ¡qué ingenioso! Aquí no hay solo un objetivo, hay también una creencia acerca de una expectativa, una hipótesis sobre la racionalidad del depredador y un plan basado en esta hipótesis; pero es extremadamente improbable que nuestro «embustero» emplumado pueda hacer tales representaciones mentales. Un monólogo más realista que represente lo que hay en la mente del pájaro sería más bien así: «He aquí a un depredador que se acerca; de repente, siento unas ansias tremendas de hacer esta absurda danza del ala rota. Me pregunto por qué…» Pero incluso este soliloquio otorga al pájaro más capacidad de reflexión de la que merece. Igual que el ascensor, el pájaro ha sido diseñado para realizar distinciones relevantes y hacer lo correcto en el momento adecuado. Los primeros investigadores, considerando, con atino, que el monólogo es demasiado bueno como para ser una descripción adecuada del «pensamiento» del ave, hipotetizaron que el comportamiento no era en absoluto deliberado, sino una especie de crisis de angustia, formada por espasmos sin dirección con el ventajoso efecto colateral de atraer la atención del depredador. Sin embargo, esta línea de razonamiento subestima muchísimo lo que el pájaro puede captar de la situación. Unos ingeniosos experimentos con los chorlitejos silbadores llevados a cabo por Ristau (1983, 1991), usando un cochecito de juguete por control remoto y que llevaba encima un mapache de peluche, mostraron que el chorlitejo vigila muy de cerca en qué está centrada la atención del depredador (la dirección de su mirada) y modula el fingimiento de su herida, aumentando de intensidad para que el depredador se acerque si este muestra signos de abandonar la caza. Por supuesto, el pájaro sale volando en el momento oportuno en cuanto el depredador ya está a una cierta distancia del nido. El chorlitejo no necesita comprender los motivos, pero identifica y responde adecuadamente a algunas de las condiciones establecidas en esos motivos. Este comportamiento no es ni un simple reflejo instintivo heredado de sus antepasados ni un astuto plan concebido en su mente racional; es una rutina diseñada por la evolución con variables que responden a diversos detalles presentes en la situación concreta, detalles que el elaborado monólogo captura (sin excesos) en la motivación de este diseño.

La motivación flotante responde a la pregunta de la ingeniería inversa «¿por qué esta rutina está organizada de la manera en que lo está?» Si somos algo remilgados con el antropomorfismo, podemos disimular y plantear la respuesta en unos términos menos «mentalistas» abusando un poco de las comillas: la rutina es un comportamiento para «captar la atención» que para su éxito depende de los «objetivos» y «percepciones» probables de un depredador, diseñada para «provocar» a este depredador a que se acerque al chorlitejo y, así, se aleje del nido; «vigilando» la «atención» del depredador y modulando el comportamiento para mantener su «interés»; el chorlitejo suele triunfar y evita la depredación de sus polluelos (esta prolija respuesta solo es superficialmente más «científica» que la versión del enfoque intencional expresado en el monólogo; ambas explicaciones se basan en las mismas distinciones, las mismas suposiciones de optimización y las mismas exigencias de información). Más resultados empíricos pueden revelar respuestas adecuadas adicionales o, por el contrario, mostrar puntos débiles en esta estratagema improvisada. Hay pruebas de que los chorlitejos saben perfectamente que no deben iniciar su engaño cuando se acerca una vaca, sino que vuelan hacia ella y la alejan del nido, en lugar de atraerla. Podemos preguntarnos si un chorlitejo puede resistir las ansias de iniciar su engaño al ver que otro pájaro, realmente herido, u otra presa vulnerable ya ha captado la atención del depredador. O también, como plantea David Haig (2014, correspondencia personal) en una sugerencia fascinante:

Podríamos imaginar un pájaro con un ala realmente rota que intenta, de modo poco convincente, escapar con la intención de que el depredador interprete sus acciones como «esto es un engaño de ala rota, de modo que en realidad el pájaro no es una presa fácil y hay un nido cerca». Si el depredador empezara a buscar el nido, habría identificado las acciones del pájaro como un mensaje pero habría malinterpretado sus motivos. La interpretación del mensaje es errónea para el depredador, pero correcta para el pájaro; el mensaje ha logrado la intención del pájaro pero ha frustrado la del depredador, que ha sido confundido deliberadamente.

Haig habla imprudentemente de los motivos e intenciones del pájaro y de la «interpretación» que el depredador hace del «mensaje», admitiendo que la tarea de pensar en estas diversas oportunidades de más experimentos y

observaciones se basan en adoptar el enfoque intencional, pero también asume que hay un grácil y gradual término medio entre interpretar a los animales (o a las plantas, los robots o los ordenadores) como si albergaran ellos mismos las razones y el razonamiento y relegar las motivaciones a la madre naturaleza como motivaciones flotantes exhibidas por la búsqueda ciega de diseños que es la selección natural.

El sentido de la señal de fingimiento de una herida tendrá su efecto previsto solo si el depredador no la identifica como una señal, sino que la interpreta como un comportamiento involuntario, y esto es cierto tanto si pájaro y depredador comprenden la situación igual que nosotros como si no. Es el riesgo de que el depredador capte lo que está pasando lo que crea la presión selectiva para un engaño más convincente del ave embustera. Igualmente, los sorprendentes y realistas «ojos» de las alas de las mariposas (los ocelos) deben su verosimilitud a la agudeza visual de sus depredadores, pero las mariposas son las beneficiarias inconscientes de su mecanismo de engaño. También hay aquí una motivación, en este caso de engaño, y decir que hay una motivación es lo mismo que decir que hay algún campo en que es predictiva y, por tanto, explicativa (para una exposición relacionada con este tema, véase Bennett 1976, párrafos 52, 53, 62). Podríamos no darnos cuenta de esto simplemente por la obviedad de lo que podemos predecir: por ejemplo, en un nicho ecológico con murciélagos como depredadores, y sin pájaros, no esperaríamos hallar polillas con ocelos (cualquier embustero racional sabe que los juegos de manos visuales sirven de bien poco con individuos ciegos o miopes).

DIFERENTES GRADOS DE COMPRENSIÓN

Ha llegado el momento de reconsiderar el lema «competencia sin comprensión». Como a menudo se asume que la competencia cognitiva es un efecto de la comprensión, me desviaré un poco para demostrar que esta suposición familiar en realidad está al revés: lo primero es la competencia. La comprensión no es el origen de la competencia ni su ingrediente activo; la comprensión está formada por competencias. Ya hemos considerado la posibilidad de otorgar una pizca, o dos, de comprensión a los sistemas que son especialmente ingeniosos en la

manera en que conducen sus competencias, pero esto nos puede llevar a una imagen engañosa de la comprensión como un fenómeno o elemento independiente que, de algún modo, se despierta por la acumulación de competencias.

La idea de la comprensión, o el entendimiento, como una maravilla mental separada e independiente es antigua pero obsoleta (pensemos en la res cogitans de Descartes, la Crítica de la razón pura de Kant o el Verstehen de Dilthey —que no es más que el término alemán para 'comprensión', pero como se escribe en mayúscula, igual que todos los sustantivos alemanes, al decirlo frunciendo el ceño evoca en muchas personas un baluarte contra el reduccionismo y el positivismo, una alternativa humanista a la ciencia—). La ilusión de que la comprensión es un fenómeno mental independiente y adicional (además del conjunto de competencias relevantes, incluida la metacompetencia necesaria para ejercer las otras competencias en el momento justo) queda reforzada por el efecto ¡ajá! o ¡eureka!, ese delicioso momento en que, de golpe, te das cuenta de que comprendes algo que hasta ese momento te dejaba perplejo. Es un fenómeno psicológico bien real, estudiado por los psicólogos desde hace décadas. Esta experiencia de una súbita aparición de comprensión se puede malinterpretar con facilidad como una prueba de que la comprensión es un tipo de experiencia (como si descubrir de repente que eres alérgico a los cacahuetes demostrara que las alergias son un tipo de sentimiento) y ha hecho que algunos pensadores insistan en que no puede haber verdadera comprensión sin conciencia (Searle [1992] es el más influyente). Así, si pensaras que es evidente que la conciencia, sea lo que sea, divide el universo en dos (todo lo que existe es consciente o no lo es; la conciencia no admite ninguna gradación), sería razonable que la comprensión, la verdadera comprensión, solo pueda ser disfrutada por seres conscientes. Los robots no comprenden nada, las zanahorias no comprenden nada, las bacterias no comprenden nada, las ostras… bueno, aún no lo sabemos, todo depende de si las ostras son conscientes; si no lo son, sus competencias, por muy admirables que sean, son competencias sin ningún atisbo de comprensión.

Mi recomendación es descartar esta forma de pensar. Este concepto casi mágico de la comprensión no tiene utilidad ni aplicación alguna en el mundo real. Sin embargo, la distinción entre comprensión e incomprensión es importante, y podemos rescatarla gracias al comprobado punto de vista darwiniano del gradualismo: hay grados y grados de comprensión. En un extremo tenemos las bacterias y su

pseudocomprensión de las señales de percepción de quorum a las que responde (Miller y Bassler 2001) y las computadoras con su pseudocomprensión de la instrucción SUMAR; en el otro extremo tenemos a Jane Austen y su comprensión de la interacción de las fuerzas personales y sociales en los estados emocionales de las personas y Albert Einstein y su comprensión de la relatividad. Pero incluso con el máximo grado de competencia, la comprensión no es nunca absoluta; en el dominio que una mente tiene de un concepto o tema siempre hay implicaciones sin entender y suposiciones no identificadas. Toda comprensión no es más que una pseudocomprensión, una comprensión a medias, desde algún punto de vista. En una charla que di una vez en el Fermilab (Illinois) a unos cuantos centenares de los mejores físicos del mundo, admití que solo tenía una pseudocomprensión de la famosa fórmula de Einstein:

$$E = mc^2$$

Puedo hacerle manipulaciones algebraicas simples, puedo decir a qué se refiere cada término y explicar a grandes rasgos cuál es su importancia, pero estoy seguro de que cualquier físico podría mostrar mi falta de comprensión de muchos de sus aspectos (los profesores somos buenos descubriendo cuándo nuestros alumnos solo tienen una pseudocomprensión, gracias a los exámenes). Luego pregunté cuántos de los presentes la entendían: se alzaron cientos de manos, como era de esperar, pero una persona se levantó y gritó: «¡No, no! Los físicos teóricos somos los que la entendemos, ¡los experimentales solo creen que la entienden!». No le faltaba razón. Por lo que respecta a la comprensión, dependemos de algo no muy diferente a la división del trabajo: confiamos en que los expertos tienen una comprensión profunda y «completa» de conceptos difíciles que los demás usamos a diario con una comprensión somera. De hecho, tal como veremos, esto es una de las contribuciones básicas del lenguaje a la inteligencia de nuestra especie: la capacidad de transmitir fielmente información que solo comprendemos a medias.

Nosotros, los seres humanos, somos los grandes expertos del planeta en esto de la comprensión, y cuando intentamos entender otras especies, tendemos a modelar su comprensión a partir de nuestra experiencia,

llenando las cabezas de los animales de sabias reflexiones, como si fueran personas de extrañas formas y con abrigos de piel. Este síndrome de Beatrix Potter, como lo he denominado, no se limita a la literatura infantil, aunque diría que todas las culturas de la tierra tienen cuentos populares y nanas sobre animales que hablan y piensan. Si hacemos esto es porque, en primera aproximación, funciona. El enfoque intencional sirve tanto si las motivaciones que alega son flotantes como si están representadas explícitamente en las mentes de los agentes analizados. Cuando un hijo aprende de su padre a determinar a qué está prestando atención una presa y cómo desbaratar su vigilancia, ambos están tratando al animal como un colega inteligente en un combate de ingenio. Pero el éxito del enfoque intencional no depende de si esto es una representación fiel de lo que pasa por la cabeza del animal, excepto hasta el punto en que lo que sea que esté sucediendo en el cerebro del animal tiene la competencia para detectar y responder adecuadamente a la información de su entorno.

El enfoque intencional nos proporciona las especificaciones de una mente y deja su ejecución para más tarde. Esto se ve muy claro en un ordenador para jugar al ajedrez: «Hazme un programa de ajedrez que no solamente sepa las reglas y siga las jugadas de todas las piezas, sino que también descubra oportunidades, identifique gambitos, suponga que el oponente hará jugadas inteligentes, valore correctamente las piezas y vigile posibles trampas. Cómo conseguirlo, es tu problema». Nosotros adoptamos la misma estrategia poco comprometida al tratar con un jugador de ajedrez humano. En una partida de ajedrez, apenas tenemos corazonadas (ni intentamos tenerlas) sobre los pensamientos detallados del oponente; esperamos que vea lo que hay que ver, que identifique las implicaciones de cualquier movimiento y que tenga buenos planes de respuesta a nuestros movimientos. Idealizamos los pensamientos de todo el mundo y también nuestro propio acceso a las razones, atribuyéndonos a la ligera episodios fantasma de agudo razonamiento a

posteriori. Tendemos a ver lo que decidimos hacer (una jugada, una compra, esquivar un golpe) como la acción correcta en el momento adecuado, y no tenemos problemas en explicar, a los demás y a nosotros mismos, cómo lo determinamos con antelación, pero al hacerlo, a menudo puede que simplemente estemos sacándonos de la manga una motivación flotante, juntándola, retrospectivamente, con nuestra experiencia subjetiva. Si nos preguntan «¿Por qué has hecho eso?», a menudo la respuesta más honesta es: «No lo sé, me salió así, nada más», pero

muchas veces caemos en la tentación de hacer un poco de historia whig,[5] buscando un para qué en lugar de conformarnos con un cómo es que.

Cuando nos ponemos a la tarea de modelizar las competencias a partir de las cuales está formada la comprensión, podemos distinguir cuatro grados, que se caracterizan de forma esquemática por aplicaciones sucesivas de la táctica conocida en informática como generar y probar. En el primer nivel, el más bajo, tenemos criaturas darwinianas, con unas competencias diseñadas y fijadas previamente, creadas por la I+D de la evolución por selección natural. Estas criaturas nacen «sabiendo» todo lo que «sabrán»; son talentosas, pero no pueden aprender. En cada generación se producen variaciones que se prueban ante la naturaleza y, en la siguiente ronda, los vencedores se copian en mayor proporción que el resto. A continuación, hallamos las criaturas skinnerianas, que, además de sus tendencias preprogramadas, tienen la tendencia clave de ajustar su comportamiento en función de «refuerzos»; generan nuevos comportamientos, más o menos al azar, que se prueban en su entorno: los que quedan reforzados (con un premio o con la eliminación de un estímulo aversivo, como el dolor o el hambre) es más probable que reaparezcan en el futuro en situaciones similares. Las variantes surgidas a partir de alguna desafortunada tendencia de identificar equivocadamente estímulos positivos y negativos, abandonando lo bueno y yendo a por lo malo, se eliminarán rápidamente, sin pasar a la descendencia. Esto no es más que el «condicionamiento operante», del cual, B. F. Skinner, el conductista por excelencia, señaló su semblanza con la evolución darwiniana, con la generación y comprobación de variantes durante la vida de un individuo, pero sin necesitar más comprensión que la propia selección natural (¡al diablo con el mentalismo!). La capacidad de mejorar el diseño propio mediante el condicionamiento operante es sin duda un rasgo que mejora la aptitud en muchas circunstancias, pero también tiene sus peligros, ya que el organismo debe probar las diversas opciones a ciegas (tan a ciegas como la evolución) en un mundo cruel y despiadado, con el riesgo de perecer antes de haber aprendido nada.

Un poco más arriba encontramos el siguiente escalón: las criaturas popperianas, que obtienen información sobre este mundo cruel y la mantienen a mano, para poder usarla al probar a priori comportamientos hipotéticos en un entorno hipotético, «fuera del mundo real», dejando que «las hipótesis mueran en su lugar», tal como expresó el filósofo de la ciencia Karl Popper. Al final, tienen que actuar en el mundo real, pero de este modo la primera elección ya no es al azar, sino que es la ganadora de

los diversos ensayos de prueba y error llevados a cabo en el modelo interno del mundo real. Para finalizar, tenemos las criaturas gregorianas, así bautizadas en honor de Richard Gregory, el psicólogo que subrayó el papel de los recursos de pensamiento al ofrecer a los pensadores lo que él llamó inteligencia potencial. El Umwelt de una criatura gregoriana está bien pertrechado con recursos de pensamiento, tanto abstractos como concretos: aritmética, democracia y estudios con enmascaramiento doble, y también microscopios, mapas y computadoras. Un pájaro en su jaula puede ver cada día tantas palabras como un ser humano (en los papeles de periódico que cubren el suelo de la jaula), pero esas palabras no son recursos de pensamiento en el Umwelt del pájaro.

Una criatura darwiniana está «preprogramada», es la beneficiaria de ingeniosos diseños que no tiene necesidad de entender. Podemos revelar su ceguera poniéndola ante nuevas variaciones de las situaciones en que ha sido diseñada por la evolución para apañarse: no aprende absolutamente nada y se esfuerza con impotencia. Una criatura skinneriana empieza con cierta plasticidad, ciertas opciones en un repertorio de comportamientos que, al nacer, todavía tienen un diseño incompleto; aprende explorando el mundo a base de pruebas y errores y está preprogramada para favorecer las incursiones que dan resultados que las refuerzan. No tiene ninguna necesidad de comprender por qué prefiere tales o cuales comportamientos ya comprobados; es la beneficiaria de este garfio de mejora de diseños, su propio proceso portátil de selección darwiniana. Una criatura popperiana observa antes de lanzarse a la acción, prueba las acciones candidatas y las confronta con el mundo que, de algún modo, ha almacenado en su cerebro. Esto ya tiene más aspecto de comprensión, porque el proceso selectivo es sensible a la información y tiene visión de futuro, pero la criatura popperiana no necesita comprender cómo o por qué se embarca en este proceso de comprobación a priori. La costumbre de crear modelos de futuro del mundo y usarlos para tomar decisiones y ajustar el comportamiento es una buena costumbre, sin duda, tanto si la comprendes como si no. Durante la infancia, a no ser que hayáis sido unos niños sorprendentemente introspectivos, os habréis lanzado «automáticamente» a la previsión popperiana y habréis cosechado algunas de sus ventajas mucho antes de haberos dado cuenta. Solamente con las criaturas gregorianas encontramos la introducción y el uso deliberado de recursos de pensamiento, la exploración sistemática de las posibles soluciones a los problemas y los intentos de control de alto nivel de las búsquedas

mentales. Parece ser que solamente los humanos somos criaturas gregorianas.

Aquí es donde tenemos que pulsar el botón rojo del excepcionalismo humano, con el encarnizado desacuerdo entre los románticos y los aguafiestas acerca de cuánta comprensión muestran tales especies o tales individuos (véase capítulo 1). La conclusión predominante en la actualidad, aunque aún tentativa, entre los expertos en inteligencia animal es que los animales más inteligentes no son «meras» criaturas skinnerianas, sino popperianas, que pueden llegar a entender algo de las ingeniosas acciones que les hemos observado realizar. Los córvidos (cuervos, urracas y otras aves de esta familia), los delfines y otros cetáceos y los primates (simios superiores e inferiores) son los animales salvajes más sorprendentes investigados hasta la fecha, con perros, gatos y loros destacándose dentro de los animales domesticados. Por ejemplo, todos ellos se embarcan en comportamientos exploratorios, analizando el terreno y a menudo dejando marcas para aligerar de peso su memoria, almacenando en su cabeza información local útil. No necesitan saber que este es el motivo de su comportamiento, pero se benefician de él al reducir la incertidumbre, ampliar su capacidad de previsión («mira antes de lanzarte» es el lema flotante de su diseño) y, en consecuencia, mejorar sus competencias. El hecho de que no comprendan el fundamento de su propia comprensión no es ningún impedimento para denominarlo comprensión, pues nosotros, los humanos, muchas veces también nos hallamos en un estado de ignorancia parecido acerca de cómo nos las apañamos para descubrir nuevas cosas. Este es el verdadero distintivo de la comprensión: la capacidad de aplicar las lecciones recibidas a nuevos elementos y a nuevos temas.

Algunos animales, como nosotros, tienen algo que podríamos equiparar a una especie de taller interno en el cual pueden dedicarse a un bricolaje de comprensión a partir de los diseños prefabricados con los que han nacido. Me permito sugerir que esta idea, de que cada organismo individual dispone de una instalación de mejora de diseños más potente que el método de fuerza bruta de probar, errar y recibir lo merecido, es el meollo de nuestra comprensión popular de la comprensión. No depende de ninguna suposición sobre la experiencia consciente, aunque este concepto es un complemento consabido, una ampliación ideológica del concepto básico. Poco a poco, nos estamos liberando de la costumbre de pensar de este modo, en parte gracias a Freud y su lucha por las motivaciones inconscientes y otros estados psicológicos, y también gracias a la ciencia

cognitiva y sus modelos detallados de los procesos inconscientes de inferencia perceptiva, búsqueda de memoria, comprensión del lenguaje y muchas más cosas. Ahora ya no se considera que una mente inconsciente sea una contradicción en sus términos; al contrario, parece que son las mentes conscientes las que plantean todos los problemas. El rompecabezas actual es «para qué sirve la conciencia (si es que sirve para algo)» si los procesos inconscientes son perfectamente competentes para desempeñar todas las operaciones cognitivas de percepción y control.

En resumen, animales, plantas y microorganismos están pertrechados con competencias que les permiten tratar adecuadamente con los ofrecimientos de sus entornos. Hay motivaciones flotantes para todas estas competencias, pero los organismos no tienen ninguna necesidad de apreciarlas ni comprenderlas para beneficiarse de ellas, ni tienen por qué ser conscientes de ellas. En los animales con comportamientos más complejos, la versatilidad y variabilidad que muestran puede justificar que les atribuyamos una cierta comprensión conductual, siempre y cuando no cometamos el error de pensar en la comprensión como en una especie de talento independiente, una fuente de competencia en lugar de una manifestación de esta.

En la parte II nos centraremos en la evolución de nosotros mismos, criaturas gregorianas, usuarios reflexivos de recursos de pensamiento. Este desarrollo es un salto de gigante en la competencia cognitiva y coloca a la especie humana en un emplazamiento único, pero, al igual que todos los procesos evolutivos, tiene que estar formado por una serie de pasos inesperados e involuntarios; la comprensión «plena» no es más que un invitado de última hora, que no se ha puesto a guiar el proceso hasta hace muy poco.

SEGUNDA PARTE

DE LA EVOLUCIÓN
AL DISEÑO INTELIGENTE

6

¿QUÉ ES LA INFORMACIÓN?

Como dicen los chinos, mil y una palabras valen más que una imagen.

J

OHN MCCARTHY

BIENVENIDOS A LA ERA DE LA INFORMACIÓN

Si nos dijeran que vivimos en la Era del Análisis, podríamos preguntarnos qué clase de análisis recibe tal honor: ¿el psicoanálisis? ¿el análisis químico? ¿el análisis estadístico? ¿el análisis conceptual? El término tiene muchos usos. Ahora los expertos nos dicen que vivimos en la Era de la Información, y el público parece estar de acuerdo, aunque pocos se dan cuenta de que este término también tiene usos diversos. ¿Qué Era de la Información? ¿La de los megabytes y el ancho de banda? ¿O la de los descubrimientos intelectuales y científicos, de la propaganda y la desinformación, de la educación universal y las amenazas a la privacidad? Los dos conceptos de información están estrechamente relacionados y, a menudo, resulta imposible desenmarañarlos en los debates, pero se pueden diferenciar. Pongamos algunos ejemplos:

El cerebro es un órgano de procesamiento de información.

Los humanos somos «informívoros» (George Miller, psicólogo).

«La información está en la luz» (J. J. Gibson, psicólogo ecologista).

La tarea del sistema nervioso es obtener información del entorno y usarla para modular o guiar un comportamiento satisfactorio.

Nos ahogamos en información.

Ya no podemos controlar nuestra información personal.

La tarea de la CIA es recoger información sobre nuestros enemigos.

La HUMINT, la información de espionaje y contraespionaje obtenida por agentes humanos en interacción clandestina con otros seres humanos, es mucho más importante que la información que se puede obtener mediante vigilancia por satélite y otros métodos de alta tecnología.

La teoría matemática de la información de Claude Shannon (Shannon 1948; Shannon y Weaver 1949) se considera, con razón, la estructura que fundamenta y legitima toda la cháchara que nos rodea acerca de la información, pero parte de esta cháchara se refiere a una concepción diferente de la información, una concepción que la teoría de Shannon solo trata de refilón. La teoría de Shannon se ocupa, en sus aspectos más básicos, de la relación estadística entre diferentes situaciones: ¿qué podemos deducir (en principio) sobre la situación A a partir de la observación del estado B? De algún modo, la situación A tiene que estar relacionada causalmente con la situación B, y esta relación causal puede variar en algo parecido a la intensidad. Shannon ideó una forma de medir la información, con independencia del contenido de esa información, igual que podemos medir el volumen de un líquido independientemente de qué líquido es (imaginemos a alguien alardeando de poseer muchos galones o cántaras pero que no pudiera responder a la pregunta «¿Galones de qué? ¿Pintura, vino, leche, gasolina?»). La teoría de Shannon ofrece un modo de desglosar la información en cantidades uniformes (bits, bytes, megabytes) que ha revolucionado todos los sistemas para almacenar y transmitir información. Esto demuestra la potencia de la digitalización de la información con las computadoras, pero de la misma manera que el vino puede ser valioso tanto si está embotellado en recipientes de volumen uniforme como si no, la

información (en el cerebro, por ejemplo) no tiene por qué estar digitalizada para poderse almacenar, transmitir y procesar.

Ahora que los CD, DVD y teléfonos móviles han sustituido a los discos de vinilo y a las transmisiones analógicas de radio, televisión y teléfono, vivimos en la Era Digital; la Era de la Información, en cambio, empezó hace mucho más tiempo, cuando la gente comenzó a escribir cosas, a trazar mapas y, en general, a registrar y transmitir información valiosa que no podían conservar en su cabeza con la fidelidad suficiente. También podemos situar antes el inicio de la Era de la Información, cuando las personas empezaron a hablar y, así, a transmitir las tradiciones, la historia y la mitología acumuladas. O también podemos decir que la Era de la Información empezó hace 530 millones de años, durante el Cámbrico, cuando se desarrolló la visión y esta desencadenó una carrera armamentística de innovación de órganos y comportamientos que podían responder con prontitud a la información obtenida mediante la luz. Incluso podemos afirmar que la Era de la Información se originó cuando surgió la vida: las células más simples que se reproducían sobrevivían gracias a elementos que funcionaban detectando diferencias en ellos mismos y en su entorno inmediato.

Para distanciar estos fenómenos de nuestras inquietudes actuales sobre los sistemas de codificación de información, los consideraré casos de información semántica, ya que identificamos la información que nos interesa

sobre una situación concreta especificando acerca de qué trata (sucesos, estados, objetos, personas, espías, productos…). Se han empleado otros términos, pero información semántica es el preferido. La información de que Pablo es alto es sobre Pablo y sobre su altura; la información de que la nieve es blanca es sobre la nieve y sobre su color. Se trata de diferentes elementos de información semántica (no les llamemos bits de información, porque bit, si bien es una unidad perfectamente válida, hace referencia al otro sentido del término información, el de Shannon). Ya antes de que la escritura estuviera generalizada, las personas inventaron formas de mejorar el control de la información semántica: rimas, ritmos y tonos musicales para fijar formulaciones valiosas en la memoria, muletas mnemotécnicas que todavía usamos hoy en día, como «agente de tráfico, guardia civil» (para recordar cómo se aparean las bases del ADN: adenina con timina y guanina con citosina), «mi vecina tiene

muchas joyas, solo una no presta» (el orden de los planetas del sistema solar… aún con Plutón) o «un día vi una vaca vestida de uniforme» (para recordar los elementos en la integración por partes).

Shannon idealizó y simplificó la tarea de desplazar información semántica del punto A al punto B descomponiéndola en un emisor y un receptor (dos agentes racionales) con un canal entre ambos y un código preestablecido y acordado, que forma el alfabeto o conjunto de señales aceptables. El canal era susceptible al ruido (cualquier cosa que interfiriese con la transmisión, degradando la señal) y la tarea era lograr una transmisión fiable que superara al ruido. Algunos diseños que conseguían esto ya estaban bien establecidos cuando Shannon concibió su teoría, como el sistema de deletreo alfabético «Able, Baker, Charlie, Dog, Easy, Fox…» utilizado por las fuerzas armadas de los Estados Unidos en las transmisiones de radiotelefonía (hasta que fue sustituido por el sistema de deletreo de la OTAN «Alfa, Bravo, Charlie, Delta, Echo, Foxtrot…» en 1955),*** para minimizar las equivocaciones entre letras que se pueden confundir con facilidad cuando se deletrea directamente con sus nombres (en inglés).

Al convertir todos los códigos, incluidas las palabras del lenguaje ordinario, a un código binario (con un alfabeto formado únicamente por dos símbolos: el 0 y el 1), Shannon demostró que la reducción del ruido se podía mejorar indefinidamente y que el coste asociado (en términos de codificación y descodificación y de disminución de la velocidad de transmisión) se podía medir con exactitud, en bits, que es un acrónimo de binary digit, ‘dígito binario’. Del mismo modo que en el juego Twenty Questions,**** en que solo se permiten preguntas a las que se pueda responder con sí o no, toda transmisión de información se puede desglosar en decisiones binarias: sí o no, 1 o 0; y a la cantidad de tales decisiones necesarias para recuperar el mensaje se le puede asignar una medida, en bits, de la cantidad de información (de Shannon) que contiene. «Estoy pensando en un número entre 0 y 8, ¿cuál es?» En el juego Twenty Questions, ¿cuántas preguntas tienes que hacer para asegurarte la respuesta? No son ocho («¿Es el cero? ¿Es el uno? ¿Es el dos? …») sino solamente tres: «¿Es el cuatro o más?» «¿Es el seis o más?» (o el dos, según la respuesta a la primera pregunta) «¿Es el siete?» Sí, sí y sí = 111 = 7 en notación binaria; se necesitan tres bits para especificar cualquier número entre el 0 y el 8. Un byte son ocho bits y un megabyte son ocho millones de bits, de modo que puedes enviar una

imagen en mapa de bits monocromático de 2,5 megabytes con el juego Twenty Million Questions («¿Es blanco el primer píxel?» y así sucesivamente).

La teoría de la información de Shannon es un gran avance para la civilización, porque la información semántica es tan importante para nosotros que queremos poder usarla de modo eficaz, almacenarla sin que haya pérdidas, desplazarla, transformarla, compartirla, ocultarla. Hay gran cantidad de artefactos para manipular información (teléfonos, libros, mapas, recetas) y la teoría de la información, de hecho, nació como un utensilio para estudiar las características relevantes de tales artefactos. Lo que empezó como una rama de la ingeniería, ha demostrado ser útil a físicos, biólogos y otros profesionales no especialmente interesados en los rasgos de los artefactos de información. Más adelante nos ocuparemos levemente de algunas de estas aplicaciones de la información de Shannon, pero nuestro objetivo principal es la información semántica.[1]

Los trucos mnemotécnicos de Shannon y sus derivados no solamente sirven para enviar información de un agente a otro, sino también para «enviar» información de un agente ahora al mismo agente en el futuro. La memoria se puede juzgar como un canal de información, sometida al ruido del mismo modo que una línea telefónica. En principio, la digitalización funciona con un alfabeto de tres, de cuatro, de diecisiete o de un millón de señales distintas, e incluso de otras formas (como veremos en otro capítulo), pero el código binario resulta ser superior para la mayoría de finalidades y siempre se puede usar; cualquier cosa se puede codificar (no perfectamente, pero sí con el grado de perfección deseado) con una estructura de ceros y unos o con una estructura de ceros, unos y doses, o… pero el código binario es el más simple de aplicar en términos físicos (encendido/apagado, tensión alta/tensión baja, derecha/izquierda) y por ello ha quedado ya casi fijado en la tecnología humana, si bien hay competencia entre códigos secundarios formados a su vez por código binario (el código ASCII para caracteres imprimibles ha quedado superado por el UTF-8, que incluye el ASCII como subconjunto; el HTML, el código empleado para las páginas web, tiene dos códigos diferentes de color en uso, el HEX y el RGB, por ejemplo).

Actualmente, la tarea de traducir, o transducir, la luz, el sonido y otros sucesos físicos que transportan información semántica al formato de

cadenas binarias de bits (el formato que se puede medir directamente como información de Shannon) es una tecnología bien asentada, y hay muchos tipos de convertidores analógico-digitales (digitalizadores) que reciben una variación continua o analógica de algún suceso físico (una onda acústica que llega a un micrófono, luz que incide en un píxel de una cámara digital, un cambio de temperatura, aceleración, humedad, pH, presión arterial, etc.) y la convierten en una cadena de bits. Estos dispositivos son semejantes a las células sensoriales que efectúan el proceso de transducción en los extremos del sistema nervioso: conos y bastoncillos de los ojos, células ciliadas de los oídos, células termosensibles, nocirreceptores (del daño), sensores de estiramiento en los músculos y, en general, todas las células de vigilancia interna que informan al sistema nervioso, incluido el sistema neurovegetativo.

En el cerebro, la conversión no es a cadenas de bits, sino a secuencias de impulsos nerviosos, diferencias de voltaje que se desplazan de una neurona a otra con aparente lentitud (millones de veces más despacio que las cadenas de bits en las computadoras). En 1943, antes de que hubiera ninguna computadora digital operativa, el neurocientífico Warren McCulloch y el lógico Walter Pitts propusieron un modo de funcionamiento de estas señales neuronales. Parecía que cuando una secuencia de impulsos nerviosos de una neurona llegaba a otra, el efecto era excitador (¡sí!) o inhibidor (¡no!); de este modo, si la neurona receptora dispusiera de algún mecanismo de umbral que pudiera sumar todos los síes y restar los noes y, luego, desencadenar su propia señal en función del resultado, podría actuar como una función lógica simple (es decir, podría actuar como una puerta AND, una puerta OR o una puerta NOT, para citar los casos más sencillos). Si algún elemento en su historia de entradas y salidas pudiera hacer aumentar o disminuir el umbral de la célula, la neurona podría «aprender» cosas que modificarían su comportamiento local. McCulloch y Pitts demostraron que una red de tales unidades podría conectarse o «entrenarse» para representar cualquier proposición, a partir de las operaciones lógicas efectuadas sobre las señales de entrada.

Fue una idea estimulante, una de las grandes simplificaciones de la historia, puesto que, si bien ha resultado que la interacción entre neuronas reales es mucho más compleja que la que las «neuronas lógicas» de McCulloch y Pitts presuponían, su hipótesis expuso la posibilidad lógica de una red genérica que pudiera representar, aprender

y controlar, formada por unidades que efectuaban tareas ciegas, sencillas y en absoluto milagrosas; en otras palabras, un «comprendedor» formado por partes que solo son competentes. Desde entonces, el objetivo de la neurociencia computacional ha sido determinar, de entre la infinita variedad de redes más complejas, cuál es la que

actúa en el sistema nervioso. Actualmente, el diagrama de conexiones del nematodo Caenorhabditis elegans, con sus 302 neuronas de 118 tipos, está casi completo y empezamos a entender su funcionamiento a la escala de impulsos individuales de neurona a neurona. En Estados Unidos, el Proyecto Conectoma Humano aspira a elaborar un mapa con el mismo grado de detalle de las decenas de miles de millones de neuronas de nuestro cerebro, mientras que en Europa, el Proyecto Cerebro Humano pretende «simular en supercomputadoras el cerebro humano en su totalidad», pero estos megaproyectos aún están en pañales. Sin duda, el cerebro humano no es una computadora digital que ejecuta código binario, pero sí que es algún tipo de computadora. En capítulos posteriores, volveremos a abordar el tema.

Por suerte, también se han conseguido grandes progresos en la comprensión de la arquitectura computacional de nuestros cerebros a una escala menos microscópica, pero estos avances se basan en dejar para más tarde las respuestas a casi todas las preguntas acerca de los enrevesadísimos detalles de la conectividad y actividad de las neuronas individuales (y que, en cualquier caso, es posible que difieran mucho entre personas, a diferencia de la estricta uniformidad que hallamos en C. elegans). Por ejemplo, descubrir que una pequeña parte concreta del cerebro (que contiene millones de neuronas) es especialmente activa cuando observamos rostros (Kanwisher 1997, 2003) es un avance relevante. Sabemos que la actividad de las neuronas en el área fusiforme facial, de algún modo, procesa la información sobre los rostros, aunque apenas tenemos idea de qué es lo que hacen estas neuronas ni de cómo lo hacen. Este uso del término información, que es habitual en las ciencias cognitivas (y en muchos otros campos), no se refiere a la información de Shannon. Hasta que se proponga una estructura de codificación (no necesariamente binaria) que digitalice el conjunto de posibilidades, no hay ningún fundamento para diferenciar la señal del ruido, ni ningún modo de medir la cantidad de información. En el futuro quizá descubramos que hay una interpretación natural de la transmisión en el sistema nervioso que nos da una medida del ancho de banda o la

capacidad de almacenamiento, en bits, de una codificación de lo que se está transmitiendo, procesando y almacenando, pero hasta entonces, el concepto de información que empleamos en las ciencias cognitivas es el de la información semántica, es decir, información acerca de algo concreto: rostros, lugares o glucosa, por ejemplo.

Si lo expresamos de otro modo, podemos decir que hoy en día los expertos en ciencias cognitivas se encuentran, más o menos, en la misma posición en que se encontraban los biólogos evolutivos y los genetistas antes del análisis de la estructura del ADN: sabían que la información acerca de rasgos fenotípicos (formas de las partes del cuerpo, comportamientos, etc.) se transmitía de generación en generación (mediante los «genes», fueran lo que fueran), pero no disponían del código ACGT de la doble hélice para proporcionarles una medida de Shannon sobre cuánta información podían pasar los padres a la descendencia en forma de herencia genética. Algunos pensadores, quizá inspirados por el ADN, creen que tiene que haber una codificación en el sistema nervioso, como el código del ADN, pero nunca he escuchado ningún argumento convincente que respaldara esta suposición y, tal como veremos pronto, hay buenas razones para ser escépticos. La información semántica, el concepto de información con el que debemos empezar, es muy independiente de la codificación, en el sentido de que dos o más observadores pueden adquirir la misma información semántica a partir de encuentros sin ningún canal compartido.[2] He aquí un ejemplo algo rebuscado:

Jacques mata a su tío en Trafalgar Square y Sherlock le arresta de inmediato. Tom lee sobre lo sucedido en The Guardian, y Boris se informa del hecho en Pravda. Jacques, Sherlock, Tom y Boris tienen experiencias muy diferentes, pero hay una cosa que comparten: información semántica acerca de que un francés ha cometido un asesinato en Trafalgar Square. No todos lo han dicho, ni tan siquiera se lo han dicho «a ellos mismos»; esta proposición, suponemos, no se les «pasó por la cabeza» a ninguno de ellos, e incluso aunque así fuera, tendría una trascendencia muy diferente para Jacques, Sherlock, Tom y Boris. No comparten ninguna codificación, pero sí comparten la información semántica.

¿CÓMO PODEMOS CARACTERIZAR LA INFORMACIÓN SEMÁNTICA?

Seguramente, la omnipresencia de los convertidores analógico-digitales en nuestras vidas, que ahora ya damos por sentada en casi todas las transmisiones de información, tiene un papel muy relevante al enmarañar los dos conceptos de información: el concepto matemático de Shannon y nuestro concepto cotidiano de información semántica, a pesar de abundantes y repetidas advertencias. Una fotografía de alta resolución y en color de un montón de confeti en una acera, dividida en ocho millones de píxeles, puede ocupar un archivo diez veces más grande que un archivo de texto que contenga, por ejemplo, La riqueza de las naciones de Adam Smith, que puede estrujarse en dos megabytes. Según el sistema de codificación que usemos (GIF, JPEG…; Word, PDF…) una imagen «valdrá más que mil palabras», si medimos este valor en bits, pero hay otro sentido en que una imagen puede valer más que mil palabras. ¿Podemos formalizar este sentido? ¿Podemos cuantificar y definir la información semántica y teorizar sobre ella? Robert Anton Wilson, novelista de ciencia ficción y periodista científico, propuso la unidad de medida jesús, definida como la cantidad de información (científica) conocida durante la vida de Jesús de Nazaret (la información científica es un subconjunto de la información semántica, que deja de lado toda la información semántica disponible en ese momento acerca de quién vivía en cada lugar, de qué color eran los vestidos de tal y cual, qué desayunaba Poncio Pilato, etc.). Por definición, en el año 30 d. C. había exactamente un jesús de información científica, una cantidad que no se duplicó (según Wilson) hasta el Renacimiento, 1500 años después. Hacia 1750, se duplicó otra vez para llegar a 4 jesuses, y a 8 jesuses en 1900; en 1964 ya había 64 jesuses, y dios sabe cuántos más se habrán acumulado desde entonces. Esta unidad no ha tenido éxito, afortunadamente, y si bien Wilson tiene buenas razones para dramatizar el tema de la explosión de información, no está claro en absoluto que ninguna escala de medida de la cantidad de información (científica) sea una mejora respecto a medidas exactas aunque tangenciales, como el número de páginas de revistas científicas con evaluadores externos o el número de megabytes de texto y datos de las revistas en línea.

Luciano Floridi, en su útil introducción (2010), caracteriza la información económica como cualquier cosa que merezca un trabajo. Para un granjero es sensato esforzarse en contar sus vacas, comprobar el nivel de agua del pozo y supervisar la eficiencia del trabajo de sus peones; si no quiere efectuar él mismo este trabajo de vigilancia, deberá contratar a alguien para que lo haga. «Compensa» saber detalles sobre tus productos, tus materias primas, tus competidores, tu capital, tu ubicación… para citar las categorías más evidentes. Podemos sondear la información pública sobre mercados y tendencias, adquirir productos de los competidores para aplicarles ingeniería inversa y compararlos con nuestros propios productos o embarcarnos en espionaje industrial. Los secretos comerciales son una categoría legal de información bien establecida, que se puede robar (o se puede filtrar o revelar irresponsablemente) y las leyes de patentes y derechos de autor estipulan restricciones al uso que otros pueden hacer de la información elaborada por la I+D de individuos y sistemas más grandes. La información económica es valiosa, a veces muy valiosa, y los métodos que se han elaborado para conservarla y protegerla de fisgones de la competencia son un reflejo de los métodos que han evolucionado en la naturaleza para el mismo propósito.

La teoría matemática de juegos (von Neumann y Morgenstern 1944) fue otra brillante innovación de la guerra. Por primera vez, se subrayaba el valor esencial de mantener en secreto las propias intenciones y planes, a salvo de los oponentes. Poner cara de póquer no sirve solo para jugar al póquer; una excesiva transparencia es una verdadera sentencia de muerte para cualquier individuo u organización que tenga que competir en el mundo cruel. En pocas palabras, la supervivencia depende de la información y, además, depende de la información diferencial o asimétrica: yo sé cosas que tú no sabes y tú sabes cosas que yo no sé; es más, nuestro bienestar depende de que las cosas se mantengan así. Incluso las bacterias (y los virus, que no están vivos) se embarcan en astutas tretas para esconderse o camuflarse, siguiendo una carrera armamentística en competición con sus intrusivos e inquisitivos competidores.[3]

Así pues, a modo de tentativa, definamos información semántica como «diseño que vale la pena obtener» y, de momento, dejemos el término diseño tan indefinido como podamos, remarcando solo el punto que he destacado en el capítulo 3, que el diseño sin diseñador (en el sentido de

diseñador inteligente) es una categoría real e importante. El diseño siempre implica una I+D de un tipo u otro y ahora ya podemos decir qué trabajo es: se trata de usar la información semántica disponible para mejorar las perspectivas de algo ajustando sus partes en la forma adecuada, sea cual sea (un organismo puede mejorar sus perspectivas adquiriendo energía o materiales —comida, medicinas, un caparazón nuevo—, pero esto no son mejoras de diseño; más bien son solo casos de reforma o reabastecimiento del propio diseño, ya existente).[4] En realidad, se puede mejorar el propio diseño como agente del mundo sin más que aprender hechos útiles (sobre dónde se pesca mejor, sobre quiénes son tus amigos). Aprender cómo se hace un anzuelo, cómo evitar a los enemigos o cómo silbar es otro tipo de mejora de diseño. Cualquier aprendizaje (tanto del qué como del cómo) puede ser un valioso complemento o actualización del diseño con el que has nacido.

A veces, la información puede ser una carga, una adquisición que interfiere con la ejecución óptima del diseño que ya se tiene; en tales casos, hemos aprendido a protegernos del conocimiento no deseado. Por ejemplo, en los experimentos doble ciego, nos esforzamos muchísimo para que tanto los participantes como los investigadores no sepan quiénes están sometidos a tal o cual situación, de modo que los sesgos en el comportamiento de los primeros y en las observaciones de los segundos sean casi imposibles. Es un ejemplo exquisito del uso de la potencia de nuestro conocimiento reflexivo, afanosamente adquirido, para mejorar nuestras capacidades futuras de obtención de conocimiento: hemos descubierto ciertos límites de nuestra propia racionalidad (como nuestra incapacidad, en determinadas circunstancias, de evitar vernos influidos por demasiada información) y usamos este conocimiento para elaborar sistemas que corrigen tal defecto. Un ejemplo más tajante, aunque menos frecuente, de conocimiento no deseado es el que se impone a los pelotones de fusilamiento: la política es repartir, al azar, un cartucho de fogueo (o un cartucho con una bala falsa de cera) entre los integrantes del pelotón, e informarles de que así se hace, de modo que ninguno pueda saber si su acción realmente causó la muerte (no se dice tan a menudo que dar cartuchos de fogueo a algunos integrantes elimina también una opción que podría resultar tentadora en algunas ocasiones: rebelarse y apuntar al oficial al mando; si sabes que has cargado un cartucho real, esta posibilidad estaría disponible como una elección racional informada; en cambio, no puedes usar información de la que no dispones).

¿Qué sucede con la información errónea, que puede acumularse, y con la desinformación, que puede inculcarse en una persona? A primera vista, estos fenómenos pueden parecer contraejemplos a nuestra propuesta de definición, pero, tal como veremos, en realidad no es así. Las definiciones de información semántica y de diseño están conectadas circularmente, pero se trata de un círculo virtuoso, no de un círculo vicioso: algunos procesos se pueden considerar procesos de I+D en los que, de cierta manera, se selecciona algún aspecto del entorno y luego se saca provecho de él para mejorar el diseño de algún conjunto o sistema de cosas relevantes, en el sentido de equiparlo mejor para medrar/persistir/reproducirse en el futuro.

La información semántica es, pues, «una distinción que marca la diferencia». Floridi (2010) afirma que fue D. M. MacKay quien acuñó esta frase por primera vez, que luego fue formulada por Gregory Bateson (1973, 1980) y otros como «una diferencia que marca la diferencia». MacKay era otro de los brillantes teóricos surgidos de los grandes esfuerzos de investigación de la segunda guerra mundial, junto a Turing y Shannon (y von Neumann y John MacCarthy, entre muchos otros). Fue un pionero de la teoría de la información, físico, neurocientífico y también filósofo (y uno de mis héroes personales, a pesar de sus profundas convicciones religiosas).[5] Mi propuesta de definición de información semántica no está muy lejos de su definición de 1950 de «información en general como lo que justifica la actividad representacional» (MacKay 1968, p. 158). En su momento, el objetivo de MacKay se centraba en lo que ahora denominamos información de Shannon, pero aventuró reflexiones sensatas sobre la información en el sentido más fundamental, más semántico, definiéndola como «lo que determina la forma» (1950, diagrama, p. 159 de MacKay 1968), lo que nos aleja de la representación en un sentido restringido a la vez que conserva la idea de justificación o valor.

Si la información es una distinción que marca la diferencia, podemos preguntarnos para quién marca esta diferencia. ¿Cui bono? ¿A quién beneficia? Esta es la pregunta que debería estar siempre en la punta de la lengua de cualquier adaptacionista, ya que la respuesta suele ser sorprendente. Es esto lo que ata la información económica de nuestra cotidianidad con la información biológica y las reúne bajo el paraguas de la información semántica, y es también lo que nos permite caracterizar la información errónea y la desinformación no como tipos de información a

secas, sino como tipos de información dependiente e incluso parasitaria. Una información aparece como errónea solamente en el contexto de un sistema diseñado para proporcionar, y basarse en, información útil. Un organismo que simplemente ignora distinciones que podrían confundir a otro organismo (dañar su diseño) no ha recibido información errónea, incluso aunque la distinción quede registrada en el sistema nervioso del organismo, si bien de algún modo poco eficaz. En el poema de Stevie Smith «Not Waving but Drowning» (1972), los curiosos en la playa que respondían al «saludo» estaban recibiendo una información errónea, pero no así las gaviotas que daban vueltas en el cielo. No podemos recibir información errónea a causa de distinciones que no estamos capacitados para hacer. Para comprender qué es la desinformación es muy beneficioso preguntarse ¿cui bono? La desinformación es la explotación diseñada (en beneficio de un agente) de los sistemas de discriminación de otro agente, diseñados para obtener información útil y usarla. Esto es lo que hace del diseño del virus del Ébola un ejemplo de camuflaje.

Kim Sterelny (2015, correspondencia personal) ha planteado una objeción relevante:

Los humanos son urracas de las representaciones (pensemos en lo rica que es la flora y la fauna de los cazadores-recolectores); buena parte de su información sobre historia natural no tiene ningún valor práctico. En cuanto el almacenamiento de información (en la cabeza o fuera de ella) se hizo barato, el hábito de la urraca se volvió adaptativo, porque es muy difícil saber de antemano cuál de los cachivaches de información resultará valioso. Pero esto no cambia el hecho de que la mayoría de ellos no lo será nunca.

Luego afirma que la mayor parte de lo que cualquier persona sabe es «adaptativamente neutro. Pero no importa, porque cuesta muy poco almacenarlo y los pequeños fragmentos que importan, importan de verdad». A grandes rasgos, estoy de acuerdo, pero me gustaría plantear una salvedad: también la basura que cada día se nos queda grabada en la cabeza con el bombardeo informativo habitual tiene su utilidad; en general, se nos queda grabada porque está diseñada para que se quede grabada, diseñada por publicistas, propagandistas y otros agentes cuyos

intereses se cumplen creando avanzadillas de reconocimiento en las mentes de otros agentes. Además, como apunta Sterelny, buena parte del resto también se nos queda grabado porque tiene una cierta probabilidad, no nula, de que pueda resultar adaptativo en algún momento (de acuerdo con nuestras evaluaciones inconscientes). Las personas (reales o míticas) que tienen una verdadera memoria fotográfica sufren una dolencia debilitante, que abruma sus cabezas con información inútil o aún peor.[6]

La evolución por selección natural se agita sin parar y extrae automáticamente pequeñas cantidades (que no bits) de información a partir de las interacciones entre fenotipos (organismos completos y pertrechados) y sus entornos. Esto lo consigue dejando que los mejores fenotipos reproduzcan sus genes con más frecuencia que los menos favorecidos.[7] A lo largo del tiempo, se «descubren» y perfeccionan los diseños, gracias a estos encuentros con la información. Realmente hay una I+D, los diseños mejoran porque todos ellos «deben costearse» en términos de reproducción diferencial, y las estirpes darwinianas «aprenden» nuevas tretas ajustando su forma. Así pues, están in-formados, un valioso progreso en el espacio de diseño local. Del mismo modo, las criaturas skinnerianas, popperianas y gregorianas se informan a sí mismas durante sus vidas gracias a sus encuentros con el entorno y son unos agentes cada vez más eficaces gracias a la información que ahora pueden usar para hacer todo tipo de cosas nuevas, como crear nuevas formas de informarse todavía más. Los ricos se vuelven cada vez más y más ricos, aprovechando su información para pulir la información que usan para perfeccionar la información que obtienen mediante los sistemas que diseñan para mejorar la información disponible cuando van a diseñar algo.

Este concepto de información útil es descendiente del concepto de ofrecimientos de J. J. Gibson, que hemos introducido en el capítulo 5. Me gustaría ampliar su exposición para incluir no solamente los ofrecimientos de plantas y otros entes no animales que evolucionan, sino también los artefactos de la cultura humana. Gibson afirma que «la información está en la luz» y que los animales perciben el mundo «captando» la información.[8] Pensemos en la luz solar reflejada que cae sobre el tronco de un árbol y sobre una ardilla que está trepando por él. En la luz hay la misma información potencial tanto para el árbol como para la ardilla, pero el árbol no está pertrechado (a causa de toda la I+D previa de su estirpe) para extraer tanta información de la luz como la

ardilla. El árbol, mediante la fotosíntesis, usa la energía de la luz para sintetizar azúcares y recientemente se ha visto que los árboles y otras plantas pueden responder adecuadamente a la información transportada por la luz: por ejemplo, determinar si germinar o no, si interrumpir la dormición, si perder las hojas o si florecer.[9]

Podemos reflexionar un poco acerca de la utilidad potencial. Supongamos que se acerca un hombre con una motosierra, visible para cualquier organismo con ojos pero no para el árbol. Los ojos no tienen utilidad alguna para el árbol, a no ser que disponga de alguna manera de usar la información (quizá no para huir y esconderse pero sí para hacer caer una rama pesada sobre el leñador o para segregar alguna savia pegajosa que trabe la sierra). La presencia de la información en la luz podría «motivar» algún día una evolución en la dirección de generar visión en los árboles, siempre que no estuviera muy lejos de alguna recompensa conductual. Es poco probable, pero estas confluencias improbables son la base de la evolución. Tiene que haber una diferencia que pudiese marcar la diferencia, solo si hubiera algo que, de algún modo, dependiera de ello. Tal como se ha señalado antes, la evolución por selección natural es increíblemente buena encontrando agujas en pajares, patrones casi invisibles que, cuando se aprovechan accidentalmente, proporcionan una ventaja a aquel que los ha aprovechado. Igual que el origen de la vida depende de disponer las materias primas moleculares correctas en el lugar adecuado y en el momento justo, tiene que haber materias primas en las variaciones de la población, entre las cuales, por casualidad, habrá ciertos rasgos hereditarios, hasta el momento sin función alguna (o infrautilizados, redundantes o vestigiales) y que varían con la información potencialmente útil que hay en el mundo.[10]

No todo lo que es «posible en principio» está automáticamente disponible, pero con el tiempo suficiente e incontables ciclos, es posible que aparezcan rutas casuales que lleguen a los Buenos Trucos de los alrededores, pero no siempre. Esta es la razón por la que los plausibles «cuentos de así fue» (Just So Stories) de Gould y Lewontin (1979) son solo explicaciones candidatas, que deben confirmarse. Toda hipótesis evolutiva bien confirmada (de las que hay a miles) empezó como un «cuento de así fue» en busca de pruebas que lo respaldaran. Y al igual que la mayoría de organismos mueren sin descendencia, la mayoría de «cuentos de así fue» concebidos nunca se ganan el derecho a

reproducirse. El pecado del adaptacionismo no es concebir cuentos de así fue (no puedes hacer biología evolutiva sin ellos), sino reproducir acríticamente cuentos de así fue que no se han comprobado adecuadamente.

Pongámonos ahora en una posibilidad menos fantástica que los árboles con ojos: el cambio de color de las hojas en otoño. ¿Es una adaptación de los árboles? Si es así, ¿en qué les beneficia? En general, se cree que no es una adaptación, sino un mero subproducto afuncional de los cambios químicos que se producen en las hojas caducas al morir. Al disminuir la luz solar, las hojas dejan de producir clorofila y, a medida que esta va desapareciendo, la luz es reflejada por otras sustancias presentes (carotenoides, flavonoides, antocianinas). Sin embargo, en la actualidad los seres humanos, sobre todo en Nueva Inglaterra, aprecian los espléndidos colores del follaje otoñal y fomentan, a menudo de forma inconsciente, la buena salud y la reproducción de los árboles más espectaculares, talando antes otros y, así, conservando los colores más atractivos para la siguiente estación ofreciéndoles un nuevo ciclo de reproducción. Para los árboles del norte de Nueva Inglaterra tener un follaje vistoso ya es una adaptación, aunque aún no se pueda medir directamente. Así es como empiezan las adaptaciones, imperceptibles para todo el mundo excepto para el proceso imperturbable de la selección natural. Hay diferencias considerables en el tiempo que conservan sus hojas en otoño los árboles caducifolios: en Nueva Inglaterra, las hojas de color marrón apagado de los robles son las últimas en caer, mucho después de que los resplandecientes arces hayan quedado desnudos. Un cambio químico que permita a algunas variedades de arce mantener su follaje durante más tiempo será una adaptación en cualquier entorno en que las personas ejerzan un efecto selectivo no desdeñable, a sabiendas o no. Forcemos ahora la imaginación un poco más: supongamos que esta prolongación del follaje es muy costosa en términos energéticos y solo compensa si hay gente en el entorno capaz de apreciar los colores; en esta situación, podría surgir la evolución de un detector de presencia de personas (más parecido a un rastreador de feromonas que no a un rudimentario ojo capaz de detectar personas). He aquí lo que sería un primer paso hacia la autodomesticación en la estirpe del árbol; en cuanto favorezcamos a su descendencia, en cuanto fomentemos o evitemos su reproducción, emprenderemos el camino hacia la domesticación, como hemos hecho con la palmera datilera o el aguacate. No es necesario que los primeros pasos sean resultado de una

elección consciente, deliberada, inteligente por nuestra parte (y, sin duda, tampoco por parte de los árboles); en realidad, podemos ser totalmente ajenos a las estirpes que se vuelven sinantrópicas, es decir, que evolucionan para medrar en compañía de los humanos sin que nos pertenezcan ni sin que nosotros las favorezcamos en modo alguno. Las chinches y las garrapatas son sinántropos, igual que los hierbajos, sin mencionar los billones de diminutos organismos que viven en nuestro cuerpo, siempre evitando nuestros radares y adaptados a vivir en el mismo nicho que el cuerpo humano o en sus cercanías.

En todos estos casos, la información semántica relativa a la mejor forma de adaptarse se ha ido espigando ciegamente, generación tras generación, y vale la pena señalar que no es una información codificada directamente en el sistema nervioso de los organismos (si es que lo tienen), ni tan solo en su ADN, excepto por algo así como una implicación pragmática. Los lingüistas y los filósofos del lenguaje emplean el término pragmática para referirse a los aspectos del significado que no quedan incluidos en la sintaxis ni en la semántica de las palabras, sino en las circunstancias y el contexto en que se expresan afirmaciones concretas; de hecho, por el Umwelt de una expresión.

Si entro en un apartamento y grito «¡poned agua a hervir!», habré expresado una frase imperativa en castellano, pero algunos de los presentes pueden deducir que tengo ganas de comer un poco de arroz hervido, o quizá de hacerme una infusión, mientras que otros pueden suponer que en este apartamento me siento como en casa e incluso que quizá sea su inquilino; otra persona, monolingüe de húngaro, solo podrá deducir que hablo castellano, igual que todos los demás que me hayan oído (o por lo menos, que hablo algo que le suena como castellano). Alguien que sepa realmente de qué va la cosa, sabrá inmediatamente que he decidido abrir con el vapor ese sobre cerrado, para leer a hurtadillas la carta que contiene, a pesar de que no soy su destinatario… está a punto de cometerse un crimen. La información semántica que puede cosecharse del suceso depende de la información que cada receptor haya acumulado previamente. Saber que alguien habla castellano puede ser una adición valiosa al conocimiento que uno tiene del mundo, una mejora de diseño que quizá algún día resulte beneficiosa. Saber que alguien está a punto de cometer un crimen también es una mejora valiosa para cualquiera que esté en situación de aprovecharla. Las mejoras locales de diseño que cada uno puede conseguir a partir de esta situación son muy diferentes, y

sería un grave error pensar que toda la información semántica se puede extraer observando con detalle la estructura de la señal como una onda acústica o como una frase de la lengua castellana («pe, o, ene, e, de, espacio, a, ge, u, a…»). No hay ningún código para todas estas diversas informaciones obtenidas.

Del mismo modo, el ADN de un ave de una estirpe que ha «aprendido» a elaborar nidos colgantes, al llegar la época de apareamiento no tiene secuencias de codones que describan esos nidos ni cómo construirlos paso a paso, sino secuencias de órdenes del tipo «ahora añade una lisina, ahora una treonina, ahora un triptófano…» (la receta para construir una proteína a partir de una cadena de aminoácidos) o «uno dos tres cuatro» (un fragmento de ADN no codificante usado como temporizador) o simplemente «bla bla bla bla» (un parásito genómico o algún trozo de ADN «basura»). Sea como fuere, no esperéis traducir ninguna secuencia de codones como nido, ramita, buscar o colocar; y, sin embargo, gracias a la cadena de codones concreta heredada por el ave a partir de sus progenitores y gracias también a los sistemas de desarrollo que, en pasos evolutivos anteriores, ya habían aprendido a interpretar esos codones, la pericia para construir nidos colgantes se habrá transmitido de padres a hijos. La descendencia hereda de sus progenitores una imagen manifiesta con una ontología de ofrecimientos y nace preparada para distinguir e identificar las cosas que le son más relevantes. Pensar en cuánto tiene que saber uno, en el caso del agua hirviendo, para comprender el mensaje en su totalidad es una buena manera de apreciar cuán impenetrable al análisis tiende a ser la transmisión de conocimiento a través del ADN de una generación a la siguiente.

Los lingüistas y filósofos del lenguaje han elaborado distinciones que parecen limitar un poco estas dificultades: así, en el acto del habla tenemos proposiciones expresadas (como «poned agua a hervir»), proposiciones implícitas (como «voy a usar el vapor para abrir el sobre») y proposiciones justificadas (como «habla castellano»).[11] Sin embargo, creo que deberíamos resistirnos a la tentación de imponer estas categorías lingüísticas en la transmisión de información del ADN, porque solo se pueden aplicar (tanto como lo permiten) a ratos y a posteriori. La evolución va de convertir «fallos» en «rasgos», de convertir «ruido» en «señal» y las fronteras borrosas entre estas categorías no son optativas: es en lo que se basa la oportunista indefinición de la selección natural. Esta es una de las claves de la

extraña inversión del razonamiento de Darwin: los creacionistas preguntan retóricamente: «¿De dónde procede toda la información del ADN?» y la respuesta de Darwin es bien simple: procede de la transformación del ruido en señal, a lo largo de miles de millones de años en un proceso gradual, sin propósito, sin milagros. Las innovaciones han de tener efectos que mejoren la adaptación desde buen principio, para poder establecer nuevas «codificaciones», de modo que la capacidad de que algo transmita información semántica no puede depender de sus cualificaciones previas como elemento de un código.

Por lo que veo, no habrá ninguna medida privilegiada para determinar cuánta información semántica transporta una señal concreta, tanto una señal genética procedente de los antecesores de uno como una señal ambiental procedente de la experiencia sensorial propia. Tal como Shannon admitió, la información siempre depende de lo que el receptor ya sabe, y aunque en los modelos podemos fijar las fronteras de la señal y del receptor, en la vida real la situación es mucho más porosa, con unas fronteras con el entorno menos definidas. En mi opinión, tendremos que apañarnos con una elucubración disciplinada, confiando en nuestra familiaridad cotidiana con la manera en que podemos comunicar tantas cosas a nuestros colegas humanos, diciendo cosas y haciendo cosas (sí, elucubración disciplinada; quizá no haya algoritmos para medir la información semántica, pero podemos construir todas las estructuras temporales que queramos para obtener una aproximación del contenido informacional de los temas que nos interesan, basándonos en ontologías propuestas: las ontologías que proveen los Umwelten de los organismos). Esto ya lo hacemos a diario, empleando nuestras categorías y ontologías humanas de forma aproximada y directa para afirmar cosas controladas con bastante rigurosidad acerca de qué categorías distinguen los animales, qué tareas llevan a cabo, qué es lo que temen, lo que desean, lo que evitan y lo que buscan. Por ejemplo, si preparamos una trampa para capturar un mapache astuto, nos preocuparemos de prestar atención a las diferencias que pueden marcar la diferencia para el mapache; el olor es una categoría evidente para tener en cuenta, pero también lo es disponer la trampa de modo que el animal, cuando se acerque, pueda vislumbrar una (aparente) ruta de escape independiente, además del lugar por el que ha llegado. Podemos esperar identificar químicamente los olores que deben enmascararse o disiparse para atraer al mapache hacia la trampa, pero los sellos distintivos del ofrecimiento ruta de escape independiente no se podrán reducir a una fórmula simple.

Tenemos que contenernos para no asumir lo que a muchos teóricos les gustaría: que si un organismo es competente con las categorías depredador, comestible, peligroso, hogar, madre, pareja, etc., debe de tener un «lenguaje de pensamiento» con términos para cada una de estas categorías. Pero si el ADN puede transportar información sobre cómo construir un nido sin necesidad de los términos construir y nido, ¿por qué no podría un sistema nervioso hacer algo igualmente inescrutable?[12]

Toda la evolución por selección natural es revisión de diseños y, en general, es mejora de diseños (o mantenimiento de diseños, como mínimo). La pérdida de órganos y de sus funciones también es una mejora cuando se tiene en cuenta el coste de mantenerlos. Los famosos peces ciegos que han descartado la visión se han dedicado a un recorte de costes, que cualquier ejecutivo de una empresa nos dirá que es una mejora de diseño: no adquieras y mantengas lo que no compensa. No es por accidente que los biólogos a menudo dicen que una estirpe «aprende» sus comportamientos instintivos generación tras generación, porque todo aprendizaje se puede considerar, de igual modo, como un proceso de rediseño propio y, en definitiva, como una mejora de diseño. Tanto la adquisición de pericias como de información factual la consideramos aprendizaje y, al fin y al cabo, siempre se trata de usar la base de conocimientos y competencias de que ya se dispone para ejercer un control de calidad sobre lo que se adquiere. Olvidar algo no se suele considerar aprendizaje, como tampoco se suele considerar una mejora de diseño desechar algo, pero a veces lo es (como en el caso de los peces ciegos). A veces más es menos. Aquí es pertinente recordar la distinción que se hace en la ley marítima asociada al derecho anglosajón entre los tipos de restos que se pueden recuperar de un naufragio: los objetos lanzados por la borda para aligerar la nave (jetsam) y los pecios u objetos que se encuentran flotando tras el hundimiento o que han caído inadvertidamente al mar (flotsam). La limpieza intencionada de la mente, el lanzamiento por la borda de información o de costumbres que pueden poner en peligro el bienestar propio no es un fenómeno inusual, a veces denominado desaprendizaje.[13]

La información semántica no siempre es valiosa para aquel que la transporta. No solo los hechos inútiles pueden sobrecargar a una persona, también determinados elementos de información pueden representar una carga emocional: la evolución no se preocupa en lo más mínimo por tus cargas emocionales, siempre y cuando tengas más descendencia que tus

competidores. Esto no anula la relación con la utilidad tal como se presenta en la definición de información semántica; más bien la complica (el valor de un puñado de monedas de oro no queda en entredicho por el hecho de que unos bolsillos llenos de esas monedas pueden hacer que un buen nadador se ahogue). Aun así, definir información semántica como «diseño que vale la pena obtener» parece desafiar el hecho de que buena parte de la información semántica que llega a nuestras cabezas día tras día es información que no vale la pena obtener o, incluso, es una odiosa molestia, que atasca nuestros sistemas de control y nos distrae de las tareas que tendríamos que estar haciendo. Pero podemos convertir este «fallo» de nuestra definición en un «rasgo» si nos damos cuenta de que la existencia misma de sistemas de manipulación de la información depende del valor de diseño de la información que justifica el gasto de construirlos. Una vez dispuesto, un sistema de manipulación de la información (un par de ojos o de orejas, una radio, Internet) puede ser explotado (parasitado) por ruido de varios tipos: ruido blanco aleatorio, sin ningún sentido (el rasposo ruido estático que interfiere en una radio cuando la señal es débil) e información semántica inútil o perjudicial para el receptor. Dos ejemplos evidentes son el correo basura y los correos fraudulentos en Internet, como también lo son las nubes de polvo y la tinta de los calamares (liberada voluntariamente). Para tener efecto, los elementos malintencionados dependen de la confianza que el receptor tenga en el medio. Desde Esopo sabemos que el niño que grita sin parar «¡que viene el lobo!» deja de recibir atención y pierde toda credibilidad después de un rato. El mimetismo batesiano, como el de una serpiente no venenosa que se parece a otra especie que sí es venenosa, es un tipo similar de parasitismo, ya que obtiene una ventaja sin el coste de producir el veneno; cuando los imitadores superan en número a las serpientes venenosas, la moraleja de Esopo queda bien demostrada, pues la señal engañosa pierde ya toda su fuerza.

Cualquier medio o canal de transmisión de información puede activar una carrera armamentística de engaño y detección, pero en el interior de un organismo los canales tienden a ser muy fiables. Como todas las «partes» tienen un destino común, ahogarse o nadar todas juntas, se establece una confianza (Sterelny 2003; para algunas excepciones fascinantes, consúltese Haig 2008 acerca del sellado genómico). Siempre puede haber errores, resultado del simple desgaste del sistema o de una aplicación incorrecta del sistema a entornos para los que no está bien equipado. Es por ello por lo que los delirios y las ilusiones son una

fuente tan rica de indicios en la neurociencia cognitiva, ya que ofrecen pistas acerca de en qué confía el organismo en una situación normal. A menudo se dice que el trabajo del cerebro en el proceso de percepción es filtrar, descartar e ignorar todas aquellas características que no son dignas de tener en cuenta del flujo de energía que incide sobre nuestros órganos de los sentidos; conservar y refinar las menas de información útil y dejar de lado toda la morralla. Cualquier elemento no aleatorio del flujo es un patrón real de información potencialmente útil, que alguna criatura o agente puede aprovechar para prever el futuro. Un pequeño subconjunto de los patrones reales del mundo de un agente cualquiera forma su Umwelt, el conjunto de sus ofrecimientos. Estos patrones son las «cosas» que el agente debería tener en su ontología, la cosas a las que debe prestar atención, las que debe vigilar, identificar y estudiar; el resto de patrones reales del flujo no son más que ruido, por lo que respecta a ese agente. Desde nuestro punto de vista olímpico (no somos dioses, pero sí que, cognitivamente, miramos al resto de organismos por encima de sus hombros), vemos con frecuencia que en el entorno hay información semántica muy relevante para el bienestar de criaturas que no pueden detectarla; la información está en la luz, sin duda, pero no es para ellas.

SECRETOS COMERCIALES, PATENTES, DERECHOS DE AUTOR Y LA INFLUENCIA DE BIRD EN EL BEBOP

Hasta el momento, mis afirmaciones son estas:

La información semántica es valiosa; la información errónea y la desinformación son dolencias o alteraciones parasitarias de los casos básicos.

El valor de la información semántica depende del receptor y no se puede medir de forma no arbitraria, pero sí que se puede confirmar mediante pruebas empíricas.

La cantidad de información semántica que transporta o contiene un suceso o elemento concreto tampoco se puede medir de modo útil con una determinada unidad de medida, pero se puede comparar de forma aproximada en circunstancias locales.

No es necesario que la información semántica esté codificada para transmitirse ni almacenarse.

Todas estas afirmaciones quedan más claras y reciben un buen espaldarazo cuando nos fijamos en la información «económica» humana y consideramos la forma en que las sociedades humanas han consagrado estas afirmaciones en sus leyes y prácticas. Pensemos en una situación en que robamos un secreto comercial; tu competidor, Cachivaches Digitales S. L., ha fabricado una nueva pieza, un componente interno de su potente cachirulizador de última generación, pero no puedes observarlo con detalle, ya que está dentro de una carcasa impenetrable a los rayos X y que si se abre o destruye, destruye también el componente de su interior. Un secreto bien resguardado, sin duda. Con grandes esfuerzos, logras introducir un espía en Cachivaches Digitales, que consigue hallar una pieza sin su carcasa. ¡Ya casi lo tienes! Pero ¿cómo extraes la información? Un molde de yeso, un negativo, sería fantástico, pero demasiado voluminoso para sacarlo a hurtadillas; esquemas, fotografías y planos serían muy interesantes, pero también costaría sacarlos de la empresa, pues la seguridad es muy estricta y se vigila cualquier señal de radiofrecuencia. Se podría cifrar una receta precisa, en castellano, para fabricar el aparato y ocultarla dentro de otro mensaje aparentemente banal, como un farragoso informe sobre opciones de seguros médicos y sus incontables condiciones, por ejemplo.

Otra posibilidad sería un archivo CAD-CAM: colocar el aparato en un TAC (una máquina para tomografías computarizadas) y obtener una representación tomográfica de alta resolución, lámina a lámina, que se pueda usar como receta para tu impresora 3D. Según la resolución, podría ser la solución ideal que, en el caso límite, ofrecería una receta para construir un duplicado átomo a átomo (una fantasía muy del agrado de filósofos y aficionados al teletransporte). Una ventaja de esta posibilidad extrema es que podemos decir que genera una especificación del aparato con el máximo detalle en un archivo cuyo tamaño se puede

medir en bits. Puedes enviar información «perfecta» sobre el aparato, especificando todos y cada uno de sus átomos, en un archivo que solo pesa tropecientos millones de zetabytes. Hoy, una pieza; mañana, el mundo. Esta idea generalizada, de describir exhaustivamente (¿?) todo el universo en un único mapa de bits cósmico es la base de muchas propuestas de la física, muy especulativas pero fascinantes. Y por supuesto, esto no se detiene a la escala de los ya entrañables átomos, hoy en día una receta «de baja resolución» para la realidad. Esta aplicación de la teoría de la información de Shannon permite, en principio, pero ni por asomo en la práctica, especificar exactamente cuánta información (de Shannon) hay, por ejemplo, en el metro cúbico de mar y de suelo marítimo que rodea a una almeja concreta, pero no dice nada sobre cuánta de esta información (una parte vastamente pequeña) es información semántica para esa almeja.[14]

Volvamos al robo del diseño de la pieza. El archivo CAD-CAM, con una resolución adecuada, se puede almacenar en una diminuta memoria digital que el espía se puede tragar sin dificultad. Pero si este no puede conseguir la tomografía, siempre puede estudiar detalladamente la pieza, observándola del derecho y del revés, sopesándola y girándola, oliéndola y lamiéndola y luego memorizar la forma y transportar la información en su cerebro (señalemos que esta es la forma en que se transmiten la mayoría de secretos: observación atenta y memorización). Quizá la mejor manera de robar la información, si puedes salirte con la tuya, es tomar prestada una pieza, llevártela a casa, examinarla y registrar la información del modo que te convenga para poder hacer una copia que cumpla tus requisitos y luego devolver la pieza a Cachivaches Digitales S. L. Todo lo que te has llevado es la información que necesitas.

Cuanto más sepa tu espía acerca de artilugios, menos información de Shannon tendrá que transmitir o sacar de Cachivaches Digitales y llevar al departamento de I+D de tu empresa. Quizá puede decir de un vistazo que la forma y el tamaño del orificio de salida es la única novedad destacable en estas circunstancias; es la única oportunidad de mejora de diseño que vale la pena espiar. Este ejemplo nos brinda una imagen más clara de la relación entre la información semántica y la información de Shannon. La idealizada restricción de Shannon de un emisor y un receptor presupone, de hecho, que el emisor ya ha encontrado la aguja en el pajar valiosa para el receptor. Buscar la aguja, detectar el patrón a aprovechar, es trabajo entre bastidores, no parte del modelo y, por tanto,

es tarea del receptor encontrar un uso adecuado para lo que recibe, aunque esta investigación y desarrollo es lo que «vale la pena» de la transmisión de la información.

Supongamos que tu espía tiene éxito. La información que has obtenido te permite mejorar el diseño de tu propio cachirulizador y, así, mejorar tu cuota de mercado y ganarte la vida. Cachivaches Digitales lo descubre (no te has preocupado de blindar tu modelo) y te demanda; o, aún peor, te encarcelan por espionaje industrial. Está más allá de cualquier duda fundada que has robado el diseño del aparato, aunque el fiscal no puede determinar cómo lo hiciste. Si la pieza es muy poco corriente, se tratará de un caso de plagio, que se puede demostrar a partir de la duplicación del diseño (con sus mejoras y variantes) y nada más. De hecho, si Cachivaches Digitales prevé la posibilidad de robo desde el principio, podrían haber añadido al diseño un elemento distintivo pero afuncional, como un botón o un mando inútil, que haría el papel de señal reveladora de robo si también aparece en tu versión (esta táctica es bien conocida en las enciclopedias para identificar a los competidores que copian ilícitamente sus entradas; el culpable no podrá explicar fácilmente que un animal, poeta o montaña ficticio aparezca en ambas obras con la misma descripción… buscad «Virginia Mountweazel» en Google). Vale la pena subrayar que esta característica reveladora solo transporta información de utilidad para el emisor original siempre y cuando el receptor original no la reconozca como señal intencional (como la señal del pájaro que simula tener un ala rota), mientras que el copiador, sin ser consciente de ello, reenvía un mensaje que le perjudica, ya que transporta la información que indica que ha copiado el original de modo ilícito.

Estas regularidades y patrones estratégicos en las interacciones entre agentes dependen de qué información se copia, no solamente de cuánta se copia; así, si bien la medida de Shannon se puede aplicar como condición limitante, no puede explicar las motivaciones flotantes de las tretas y contratretas. En muchos campos de la biología es útil encontrar información reveladora en los sistemas biológicos, información que podemos concluir que tiene que estar ahí, aunque aún no tengamos un conocimiento detallado de cómo se «codifica» o se encarna tal información. Por ejemplo, mi colega Michael Levin (2014; Friston, Levin et al. 2015) ha elaborado modelos de morfogénesis que tratan «los sistemas que forman patrones como agentes cognitivos primitivos»

(sistemas intencionales simples). Las neuronas no son las únicas células con «conocimiento» y «programas» (véase el capítulo 8).

Creo que podemos aprender otra cosa a partir de la legislación de patentes y de derechos de autor. En primer lugar, estas leyes se promulgan para proteger los diseños, en este caso diseños creados por diseñadores inteligentes, es decir, personas. En muchos casos, las personas somos diseñadores, y el proceso de diseño lleva tiempo y energía (y un poco de inteligencia, a no ser que te lances a un incansable proceso de prueba y error, un método de I+D que casi nunca da frutos destacables excepto a lo largo de los periodos prolongados de la escala evolutiva). Los diseños resultantes suelen ser valiosos (para alguien), de modo que parece razonable que haya leyes que protejan a los creadores o propietarios de esos diseños.

Cabe destacar algunas particularidades de estas leyes. En primer lugar, para obtener una patente tienes que demostrar la utilidad de tu creación y también que nadie más lo ha inventado antes. Pero exactamente, ¿cuán útil y cuán original tiene que ser? Aquí es donde entra en escena la inevitable ambigüedad de la ley. Por ejemplo, la ley canadiense de patentes excluye como no suficientemente novedosa (para merecer una patente) cualquier invención para la que exista anticipación, que se puede demostrar si se cumple alguna de las ocho condiciones establecidas (véase la entrada «Novelty (patent)» en la Wikipedia en inglés); dos de estas condiciones se describen del modo siguiente: la anticipación debe

transmitir información de modo que una persona que se enfrente al mismo problema debe poder decir «esto me da lo que deseo»;

dar información a una persona con conocimientos generales de modo que se percate de inmediato de la invención.

No debería sorprendernos que las leyes de patentes tengan estos incómodos problemas de definición. La novedad, como la información semántica en general, depende muy directamente de la competencia de las partes implicadas. En Zoquetelandia podrías patentar un ladrillo como tope de puerta; en cambio, en Edén de los Ingenieros, una casa voladora alimentada por luz solar podría considerarse una ampliación trivial del conocimiento y las prácticas existentes. Lo que una «persona con conocimientos generales» puede captar «de inmediato» cambiará en función de lo que se considere conocimiento general en cada lugar y en cada momento.

Puedes patentar una idea (de un proceso, de un aparato, de un utensilio, de un método) sin construir un prototipo operativo (aunque está claro que es recomendable fabricar uno) que respalde tus esquemas y descripciones. En cambio, no puedes registrar los derechos de autor de una idea, sino solo una expresión concreta de una idea. Una canción puede estar sometida a derechos de autor, pero es muy posible que una serie de cuatro notas, no. Si Beethoven estuviera vivo, ¿podría registrar los derechos de autor de las primeras cuatro notas de su quinta sinfonía (¡ta ta ta chaaan!)? ¿Tiene la NBC derechos de autor sobre su repique de tres notas***** o es solamente una marca comercial (un elemento de información con una protección legal diferente)? El título de un libro no se puede someter a derechos de autor, pero un pequeño poema sí… ¿qué tal este?

En este poema

flaquea el tema.

Sin duda, queda cubierto por los derechos de autor de todo el libro, pero me pregunto si también lo estaría como «obra literaria» independiente. Las leyes de derechos de autor se han modificado muchas veces, pero sigue habiendo aspectos controvertidos. Los libros y los artículos, la música y la pintura, los dibujos, la escultura, la coreografía y la arquitectura se pueden someter a derechos de autor siempre y cuando

haya alguna expresión fija y concreta. Un solo de jazz improvisado en una actuación no grabada no se puede someter a derechos, como tampoco una rutina de baile sin coreografía y sin filmar. Esto encaja con la definición de Colgate y Ziock de información (véase la nota 7 de este capítulo), pero parece quedar determinado por requisitos legales probatorios, no por ningún requisito natural de que la información esté «almacenada (escrita)». Charlie 'Bird' Parker no podía registrar derechos de autor por sus solos, pero influyeron en muchos otros saxofonistas y músicos de jazz que los escucharon, lo que equivale a decir: información semántica valiosa ha fluido de las actuaciones de Bird a las de los otros (y no a la audiencia con poca sensibilidad musical, que no puede recoger los ofrecimientos que hay disponibles). Una idea o un descubrimiento no puede tener derechos de autor, pero «¿dónde trazamos la frontera?» Tal como dijo el juez Learned Hand, «está claro que no se puede afirmar ningún principio sobre cuándo un imitador ha ido más allá de copiar la "idea" y, en su lugar, ha tomado prestada su "expresión". Las decisiones deben tomarse caso por caso» (caso de Peter Pan Fabrics, Inc. contra Martin Weiner Corp., 274 F.2d 487 [2d Cir. 1960], citado en la entrada «copyright» de la Wikipedia en inglés).

Es interesante que, cuando se considera si un elemento puede someterse a derechos de autor, la utilidad o la función va en contra de una creación, ya que los derechos de autor están pensados para proteger la creación «artística», que se debe poder «separar conceptualmente» de las consideraciones funcionales, para las que es aplicable la ley de patentes (bajo condiciones más restrictivas). Se trata de una interesante y amputada idea de función, ya que los efectos estéticos son claramente funcionales en la mayoría de contextos, si no en todos. Esto me recuerda al miope rechazo que planteaban algunos de los primeros evolucionistas de las ideas de Darwin sobre selección sexual, ya que esta implicaba que la percepción de la belleza podría tener un papel funcional, algo erróneo, a su parecer… en realidad, ¡y tanto que tiene un papel funcional! Las hembras han evolucionado para tener unas competencias agudizadas en la identificación de propiedades favorables de las posibles parejas, mientras que los machos han evolucionado para hacer unas exhibiciones cada vez más impresionantes (estéticamente impresionantes, buenas solo como ostentación). Un apareamiento satisfactorio no es una opción en la vida, es la línea de llegada, el objetivo, y sea lo que sea lo que permita llegar al final, será funcional, cueste lo que cueste, a cualquier precio.[15] Las leyes de derechos de autor intentan separar la función utilitaria de la

función estética y, si bien hay buenas razones legales para que esta separación sea una línea clara y nítida, también hay buenas razones teóricas para considerarla una pretensión ad hoc, tal como expresa la distinción de Learned Hand entre idea y expresión.

Cuando pasemos a la evolución cultural, nos toparemos con esta especie de transmisión sin código a gran escala. Tal como he señalado a menudo, un carro con ruedas de radios no solo transporta grano o mercancías de un lugar a otro, transporta también la brillante idea de un carro con ruedas de radios. El carro no transporta esta idea, esta información, al perro que lo ve pasar por el camino, no más que la luz del Sol transporta al árbol la información del leñador que se acerca. Para empezar, tienes que estar informado, tienes que tener muchas competencias en tu haber, antes de poder servirte de esta información, pero sin embargo, ahí está, encarnada en el vehículo que pasa por el camino. Una de las tareas que nos quedan es comprender cómo es que nosotros, los seres humanos, somos tan buenos para extraer información del entorno, mucho más que cualquier otra especie.

Pero, ¿de verdad somos mejores? Sí, sin el menor atisbo de duda, si medimos nuestra destreza en número de ofrecimientos. Además de los que compartimos con nuestros parientes mamíferos (agua para beber, comida para alimentar, recovecos en donde esconderse, senderos que seguir…), tenemos nuestros artefactos familiares e identificables. Una ferretería es un museo de ofrecimientos, con cientos de diferentes utensilios para unir, abrir, cerrar, extender, cavar, alisar, recoger, abrazar, asir, cortar, escribir, almacenar, etc., todos ellos identificables y utilizables en las circunstancias adecuadas, incluyendo circunstancias novedosas en que improvisamos para inventar y construir nuevos ofrecimientos a partir de partes prediseñadas y prefabricadas. Richard Gregory, cuyas reflexiones acerca de la inteligencia y el equipamiento me inspiraron a bautizar las criaturas gregorianas en su honor, hizo hincapié en que no solo se necesita inteligencia para usar unas tijeras, sino que las tijeras potencian la inteligencia de sus usuarios al facilitarles un gran aumento de las competencias disponibles. Estos utensilios, como el caparazón del cangrejo ermitaño, el nido del pájaro o la presa del castor, son mejoras de diseño adquiridas, no forman parte de nuestros fenotipos extendidos (Dawkins 1982, 2004b); el rasgo fenotípico que transmiten nuestros genes es el talento general para identificar estas cosas y aprovecharlas a nuestro favor.

A veces sugiero a mis estudiantes que la evolución por selección natural no es nada más que un «plagio universal»: si te sirve de algo, cópialo y úsalo. Toda la I+D invertida en configurar lo que sea que estás copiando ahora ya es parte de tu legado, lo has añadido a tu riqueza sin tener que «reinventar la rueda». Esto es lo que la naturaleza ha estado haciendo durante miles de millones de años, perfeccionando, articulando y difundiendo miles de millones de buenos rasgos de diseño en cada rincón del planeta. Esta formidable creatividad no sería posible sin una cantidad casi inimaginable de copias. Para nosotros, el Buen Truco de la naturaleza va contra la ley y hay una buena razón para ello: la información semántica es costosa y valiosa, de modo que las copias no autorizadas equivalen a un hurto. Cabe remarcar que no se trata de una motivación flotante: las primeras leyes de patentes y derechos de autor (y de secretos y marcas comerciales) se concibieron y se promulgaron tras debates explícitos y prolongados y una discusión sobre su necesidad; son el producto de un diseño inteligente, ideado para proteger otros diseños inteligentes.

La información de Shannon nos proporciona un armazón matemático para distinguir la señal del ruido y para medir la capacidad y la fiabilidad. Esto clarifica el entorno físico en que debe tener lugar toda I+D, pero la propia I+D, el desarrollo de «dispositivos» de detección de patrones que pueden refinar las vetas, que pueden encontrar las agujas de los pajares, es un proceso que apenas ahora empezamos a comprender de abajo arriba. Hasta ahora hemos podido razonar acerca de la información semántica necesaria para diversos propósitos (comunicar las elecciones racionales, dirigir, construir una ratonera mejor, controlar un ascensor) independientemente de las consideraciones acerca de cómo estaba encarnada físicamente esta información semántica. Tal como afirmó Norbert Wiener, el padre de la cibernética (1961): «la información es información, no es materia ni energía. Ningún materialismo que no admita esto podrá sobrevivir actualmente».

*** En español, el alfabeto habitual de deletreo, a parte del de la OTAN, es Antonio, Burgos, Carmen, Chocolate, Dolores, Enrique, Francia… (N. del t.)

**** Popular juego de radio y televisión en Estados Unidos, consistente en descubrir algo haciendo un máximo de veinte preguntas a las que solo

se puede responder con un sí o con un no. Como juego de mesa, fue publicado en español en la década de 1990. (N. del t.)

***** Se refiere a la sintonía de tres notas que identifica las emisiones y programas de la NBC: sol2, mi3, do3; el lector puede oírlo aquí: https://en.wikipedia.org/wiki/File:NBC_chimes.ogg. (N. del t.)

7

UN ENTREACTO DE ESPACIOS DARWINIANOS

UN NUEVO RECURSO DE PENSAMIENTO SOBRE LA EVOLUCIÓN

Darwin, en su resumen del capítulo 4 de El origen de las especies, ofreció el marco básico para entender la evolución por selección natural. Vale la pena reproducirlo:

Si en condiciones variables de vida los seres orgánicos presentan diferencias individuales en casi todas las partes de su estructura —y esto es indiscutible—; si hay, debido a su progresión geométrica, una rigurosa lucha por la vida en alguna edad, estación o año —y esto, ciertamente, es indiscutible—; considerando entonces la complejidad infinita de las relaciones de los seres orgánicos entre sí y con sus condiciones de vida, que hacen que sea ventajoso para ellos una infinita diversidad de estructura, constitución y costumbres, sería un hecho de lo más extraordinario que no se hubiesen presentado nunca variaciones útiles a la prosperidad de cada ser, del mismo modo que se han presentado tantas variaciones útiles al hombre. Pero si las variaciones útiles a un ser orgánico ocurren alguna vez, los individuos caracterizados de este modo tendrán seguramente las mayores probabilidades de conservarse en la lucha por la vida, y, por el poderoso principio de la herencia, tenderán a producir descendientes con caracteres semejantes. A este principio de conservación o supervivencia de los más adecuados lo he llamado selección natural.

A lo largo de los años, esta gran intuición se ha perfeccionado y generalizado y se han elaborado diversas formulaciones compactas. La mejor, por su simplicidad, generalidad y claridad quizá es la del filósofo de la biología Peter Godfrey-Smith (2007):

La evolución por selección natural es el cambio en una población debido a

variación en las características de los miembros de la población,

que provoca velocidades diferentes de reproducción y

que es heredable.

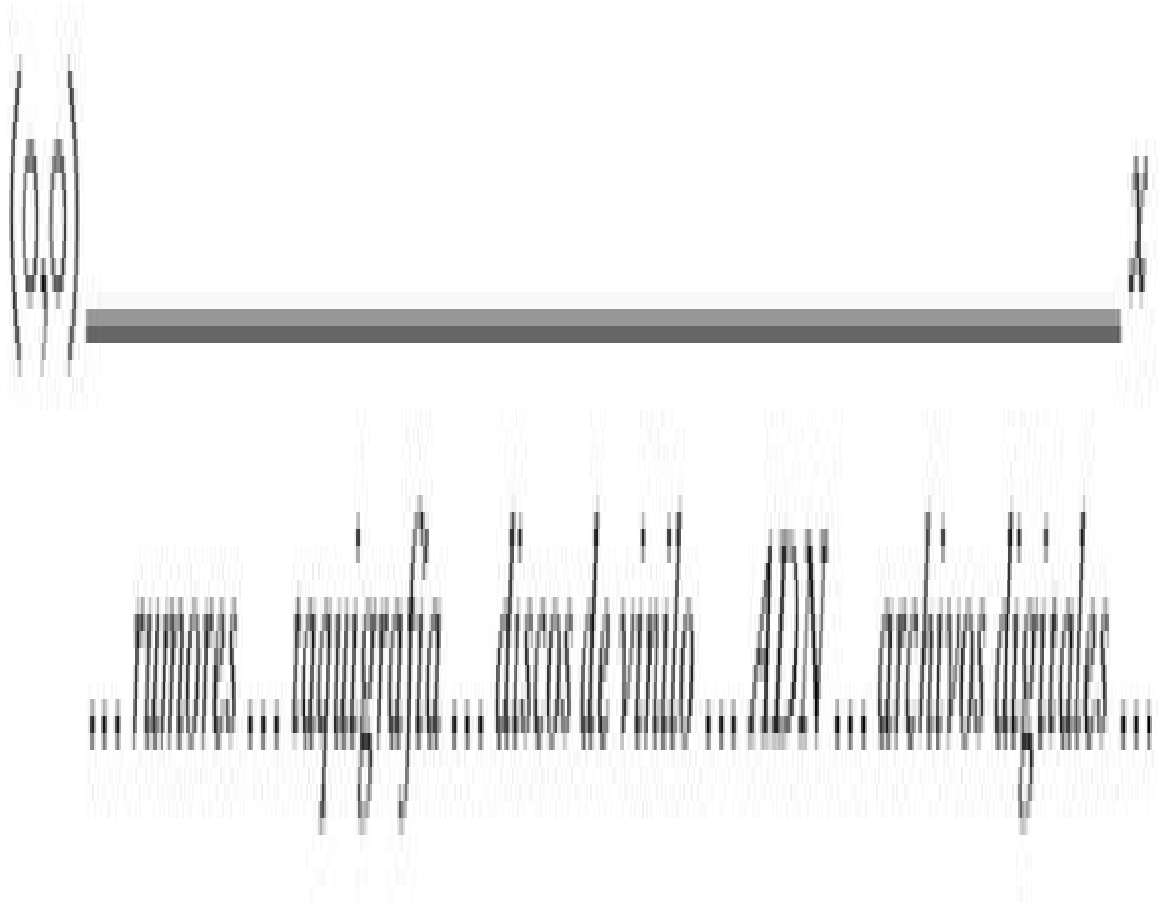
Siempre que se dan estos tres factores, el resultado inevitable es la evolución por selección natural, tanto si la población está formada por organismos como por virus, programas informáticos, palabras o cualquier otra variedad de cosas que generan copias de sí mismas de un modo u otro. Con cierto anacronismo, podemos expresar esto diciendo que Darwin descubrió el algoritmo fundamental de la evolución por selección natural, una estructura abstracta que se puede poner en práctica, puede «realizarse», en diferentes materiales o medios.

Echaremos un vistazo a la evolución de la cultura por selección natural (objetos culturales, como palabras, métodos, rituales, estilos) y para ayudar a nuestra imaginación a cavilar sobre este traicionero tema, me gustaría presentar un recurso de pensamiento que creo que es particularmente valioso para orientarnos en este terreno resbaladizo: los espacios darwinianos inventados por Godfrey-Smith en su libro Darwinian Populations and Natural Selection (2009).

Es tentador afirmar que las tres condiciones anteriores definen la esencia de la selección natural darwiniana, y los ejemplos clásicos de selección natural que se usan para ilustrar la teoría en acción encajan perfectamente. Pero una de las contribuciones más importantes de Darwin al

pensamiento fue su negación del esencialismo, esa vieja doctrina filosófica que afirmaba que para cada tipo de cosa, cada clase natural, hay una esencia, un conjunto de propiedades necesarias y suficientes para ser ese tipo de cosa. Darwin mostró que las diversas especies están conectadas históricamente por una cadena de variaciones que difieren tan gradualmente que no hay forma honesta de trazar una frontera y decir, por ejemplo: los dinosaurios a la izquierda, las aves a la derecha. Pero entonces, ¿qué sucede con el darwinismo? ¿Tiene alguna esencia, o es propenso a variaciones que se difuminan de forma imperceptible en explicaciones no darwinianas? ¿Qué sucede con los fenómenos que no son casos tan paradigmáticos de evolución darwiniana? De estos, hay muchos, y Godfrey-Smith muestra cómo organizarlos de forma que se destaquen con claridad sus diferencias y semblanzas e incluso sugieran sus propias explicaciones.

Por ejemplo, todas las entidades que evolucionan necesitan «heredar» algo, una copia de algo, de sus progenitores, pero algunos tipos de copia son de baja fidelidad, con muchas distorsiones y pérdidas mientras que otros son réplicas casi perfectas. Podemos crear una gráfica simple con un eje x para ordenar todos los casos posibles, de las copias de baja fidelidad a las de alta fidelidad:



La evolución depende de la existencia de copias de alta fidelidad, pero que no sean copias perfectas, ya que las mutaciones (los errores de copia) son la fuente definitiva de toda novedad. La tecnología digital de copia es perfecta para cualquier propósito práctico: si copias una copia de una copia de una copia… de un archivo Word, será idéntico, letra a letra, al archivo original. Para bien o para mal, no hay mutaciones que se vayan a acumular. El ADN se copia a sí mismo de forma casi perfecta, pero sin los errores ocasionales (de los cuales no se produce ni uno por cada mil millones de nucleótidos), la evolución se pararía en seco.

He aquí otro ejemplo de variación: las diferencias en aptitud entre miembros de una población pueden depender de la «suerte», del «talento» o de cualquier combinación de ambas. En un entorno selectivo en que a la mayoría de los que no se reproducen les alcanza un rayo antes de tener oportunidad de aparearse, la evolución no puede actuar sobre las eventuales diferencias de destreza que haya. Podemos ordenar las

diversas proporciones de suerte y de talento en otro eje y darnos cuenta de que, igual que el «ruido» en las copias (replicación de baja fidelidad) entorpece la I+D evolutiva, el «ruido» ambiental (rayos u otros accidentes que eliminan candidatos meritorios para el torneo de la reproducción) también lo hace. Y ya que estamos, podemos añadir una tercera dimensión y representar un cubo con ejes x, y y z, sobre el cual situar diversos fenómenos (por desgracia, la mayoría no podemos visualizar cuatro o más dimensiones con facilidad, de modo que nos conformaremos con tres y estudiaremos solo tres tipos de variación a la vez, poniendo y quitando las que nos interesen en cada momento).

Gracias a estas representaciones tridimensionales, podemos visualizar fenómenos darwinianos puros, cuasidarwinianos, protodarwinianos y (saltándonos las desdibujadas fronteras) también fenómenos en absoluto darwinianos. Poder observar de un vistazo cómo se parecen y cómo difieren entre sí todos estos fenómenos es una elegante ayuda para pensar acerca de la evolución. En lugar de intentar trazar demarcaciones claras para separar fenómenos meramente pseudodarwinianos de los fenómenos que presentan los tres rasgos esenciales de la selección natural, creamos gradientes sobre los que situar las cosas que son más o menos darwinianas en algún aspecto u otro. Así, podemos ver si la teoría darwiniana es aplicable a los casos intermedios, buscando términos medios e interacciones que puedan explicar cómo y por qué estos fenómenos se producen del modo en que lo hacen. Godfrey-Smith nos recuerda que «los procesos evolutivos mismos son productos de la evolución» y que, en consecuencia, emergen y se transforman gradualmente. En resumidas cuentas, sus espacios darwinianos nos ofrecen un recurso provechoso para ayudarnos a sostener y profundizar nuestro «darwinismo sobre el darwinismo», citando el excelente concepto de Glenn Adelson.

En todos estos diagramas, consideramos que las tres dimensiones toman valores entre 0 (en absoluto darwiniano) y 1 (totalmente darwiniano), de modo que los fenómenos paradigmáticos del «darwinismo total» se sitúan en el vértice superior derecho, con un valor de 1 en las tres dimensiones: (1,1,1); en cambio, los fenómenos que no muestran ninguna de las características representadas se sitúan en el (0,0,0). Podemos escoger tres variables cualesquiera y asignarlas a las tres dimensiones x, y y z. En la figura 7.1, la fidelidad de la herencia, H, se representa de izquierda a derecha en el eje x; si la fidelidad es demasiado baja para conservar la señal en presencia de ruido, no se producirá ningún tipo de evolución:

pueden surgir mutaciones útiles, pero desaparecerán ahogadas por el ruido antes de que la selección pueda dirigirlas hacia las generaciones sucesivas. Incluso si las otras dimensiones se hallan a valores darwinianos adecuados (cercanos a 1), obtenemos las «catástrofes de errores» que se alinean cerca del lado izquierdo (recordad que, tal como ya hemos recalcado, si la fidelidad es perfecta, la evolución también se detiene por falta de nuevas variaciones, de modo que los casos correspondientes al paradigma deben estar muy cerca del lado derecho, pero no justo en él). El eje vertical y representa la continuidad, o la «suavidad del paisaje adaptativo». La selección natural es un proceso gradual; depende de pequeños pasos a ciegas y si estos pequeños pasos se producen en una pendiente suave o uniforme (de modo que un paso a un costado en cualquier dirección también sea un paso hacia arriba, hacia abajo o neutro en la aptitud), una sucesión de pequeños pasos puede «escalar una montaña» y, a pesar de su miopía (el relojero ciego), culminar un pico. Podrá llegar al pico más alto, el óptimo global, en un paisaje adaptativo especialmente suave, pero llegará solo a un óptimo local en un paisaje con varios picos. Cuando el paisaje es «accidentado», la evolución es casi imposible, porque los pequeños pasos no se corresponden con un progreso o, aún peor, ni tan siquiera pueden mantener la aptitud actual.

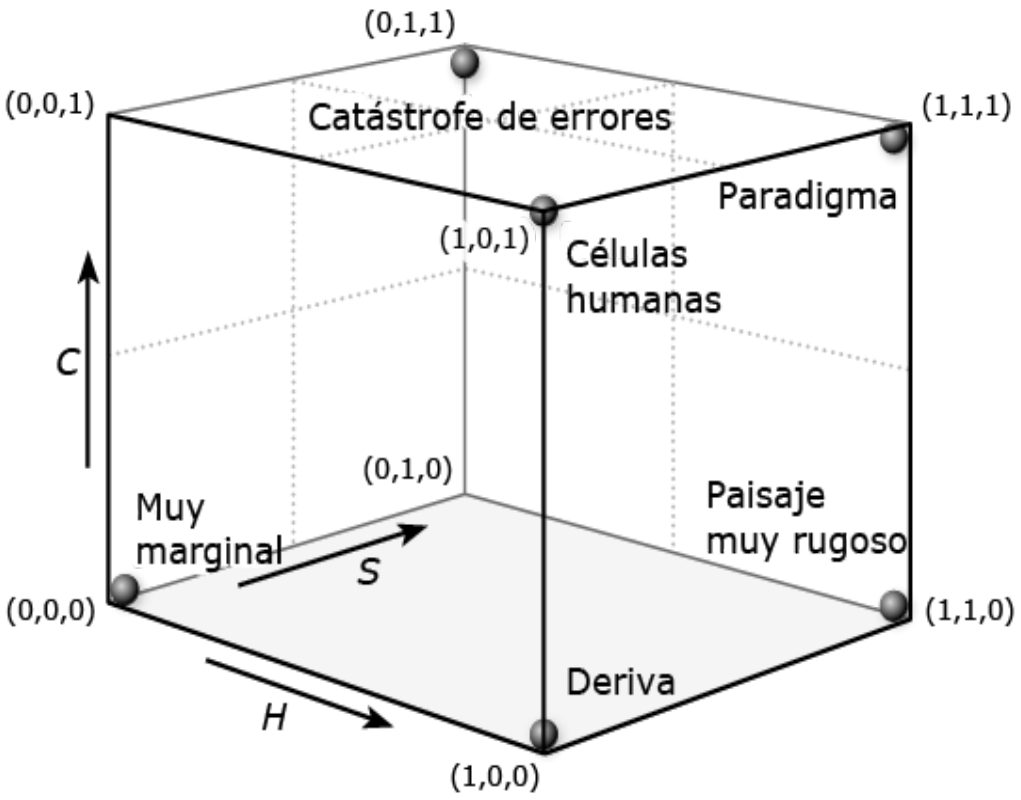


FIGURA 7.1. Espacio darwiniano. © Godfrey-Smith

En la dimensión z, Godfrey-Smith representa la dependencia de las «propiedades intrínsecas», S (no preguntéis el porqué de esta S), característica que captura la distinción entre suerte y talento. En un torneo de tenis, los mejores atletas suelen avanzar en la competición, gracias a su competencia y fortaleza; en un torneo de cara y cruz, el resultado viene determinado por la pura suerte, no por ninguna propiedad intrínseca de los participantes, y el vencedor no tiene más posibilidades de superar a los perdedores en una revancha que cualquier otro jugador. Por ejemplo, en la deriva genética (sobre todo en poblaciones pequeñas en que los «errores de muestreo» son comparativamente grandes), el rasgo «ganador» de una población no es mejor que las alternativas; simplemente la suerte se ha ido acumulando en una dirección,

potenciando un rasgo que ha quedado fijado en esa población. En una pequeña población de conejos en que unos pocos tienen el pelaje más oscuro que los otros, puede suceder que algunos de los más oscuros mueran a causa de un atropellamiento en una carretera, otro se ahogue y el último portador del gen de pelaje oscuro caiga por un precipicio justo cuando iba a aparearse; de este modo el gen de pelaje oscuro ha desaparecido del patrimonio genético, pero no por ser peor.

Una de las particularidades más valiosas de estos espacios darwinianos es que nos ayudan a ver el fenómeno de la «desdarwinización», en que una estirpe que ha ido evolucionando durante generaciones en condiciones darwinianas paradigmáticas se desplaza a un entorno nuevo en que su futuro pasa a estar determinado por un proceso menos darwiniano. Las células humanas, que se muestran en el diagrama, constituyen un buen ejemplo. Estas células son descendientes directos pero muy lejanos de los eucariotas unicelulares (microbios o protistas) que se las apañaron por sí mismos como individuos independientes. Estos lejanos ancestros eran criaturas en una evolución darwiniana paradigmática (en el vértice [1,1,1]), controlada totalmente por los tres rasgos esenciales. Las células que actualmente forman nuestro cuerpo son descendientes directos (las «células hijas») de células que, a su vez, descienden de células anteriores, hasta el zigoto que se formó cuando el óvulo y el espermatozoide se unieron en el momento de la fecundación. Durante el desarrollo, en el útero materno, y en la infancia, se produce una gran proliferación de células, muchas más de las que, al final, se necesitarán para formar los órganos; por ello, un proceso de selección descarta sin compasión los sobrantes y conserva los ganadores, que se encargarán de diversas tareas, mientras que los perdedores se reciclan como materia prima para la siguiente «generación». Este proceso es muy evidente en el cerebro, donde se da oportunidad a las neuronas nuevecitas para que se conecten de modo útil (por ejemplo, formando parte de una vía que va de un determinado punto de una de nuestras retinas hasta algún punto correspondiente en nuestra corteza visual). No es muy diferente de una carrera, con muchas neuronas que intentan crecer de A a B siguiendo pistas moleculares como si fueran miguitas de pan en el bosque; las que llegan primero ganan y sobreviven, y el resto muere y se convierte en la materia prima de la siguiente hornada.

¿Cómo saben los ganadores cómo crecer desde A hasta B? No lo saben, es pura suerte (como ganar a cara o cruz). Muchas neuronas intentan el viaje y la mayoría fracasan, mientras que las que consiguen la conexión

correcta se salvan. Las células que sobreviven pueden ser «intrínsecamente» iguales a sus competidoras (no son más fuertes ni más rápidas), simplemente están en el lugar correcto en el momento adecuado. Así pues, el proceso de desarrollo que conecta el cerebro es una versión desdarwinizada del proceso que hizo evolucionar los eucariotas que unieron sus fuerzas hace miles de millones de años para crear organismos multicelulares. Fijémonos que la deriva genética, en la cual un rasgo queda fijado como resultado de la suerte (como las células humanas), se muestra en el diagrama con un valor bajo para escalar la montaña; en el caso de las células humanas, la ubicación es fundamental, como diría un agente de la propiedad inmobiliaria, pero los ganadores de la deriva genética solamente han sido afortunados, sin más. La deriva genética ha estado presente en el mundo tanto como la evolución, de modo que no se trata de un caso de desdarwinización.

En la figura 7.2 presento otro espacio darwiniano de Godfrey-Smith. En este caso, el eje x corresponde a B, los cuellos de botella: ¿Pasa la reproducción por algún tipo de embudo (en el caso extremo, una única célula)? De los varios billones de células que forman al padre y a la madre en el momento de la concepción, se unen un único espermatozoide y un único óvulo para dar inicio a la nueva generación. En cambio, una manada de búfalos se convierte en dos manadas (que luego pueden ser cuatro) sin más que separarse en dos grupos, pero también puede pasar que una única pareja de búfalos quede aislada y sean los fundadores de una nueva manada, con mucha menos diversidad genética. Cualquier parte de una esponja se puede convertir también en una esponja descendiente, pero una esponja parece algo más integrado, una cosa más independiente, que una manada. Un roble es algo muy integrado y tiene un cuello de botella reproductivo: la bellota, pero, igual que las plantas en general, se puede reproducir a partir de un retoño plantado en el suelo. De momento, no hay forma de crear un ser humano descendiente (un clon) a partir de una loncha de una oreja o de una tajada de un dedo del pie. En las tres dimensiones mostradas nos hallamos en valores altos. Las alamedas son interesantes porque están interconectadas bajo el suelo mediante sistemas radiculares, de modo que, desde un punto vista genético, cada álamo no es un individuo sino que forman algo así como unos megagemelos siameses, una única «planta» que cubre toda una colina.

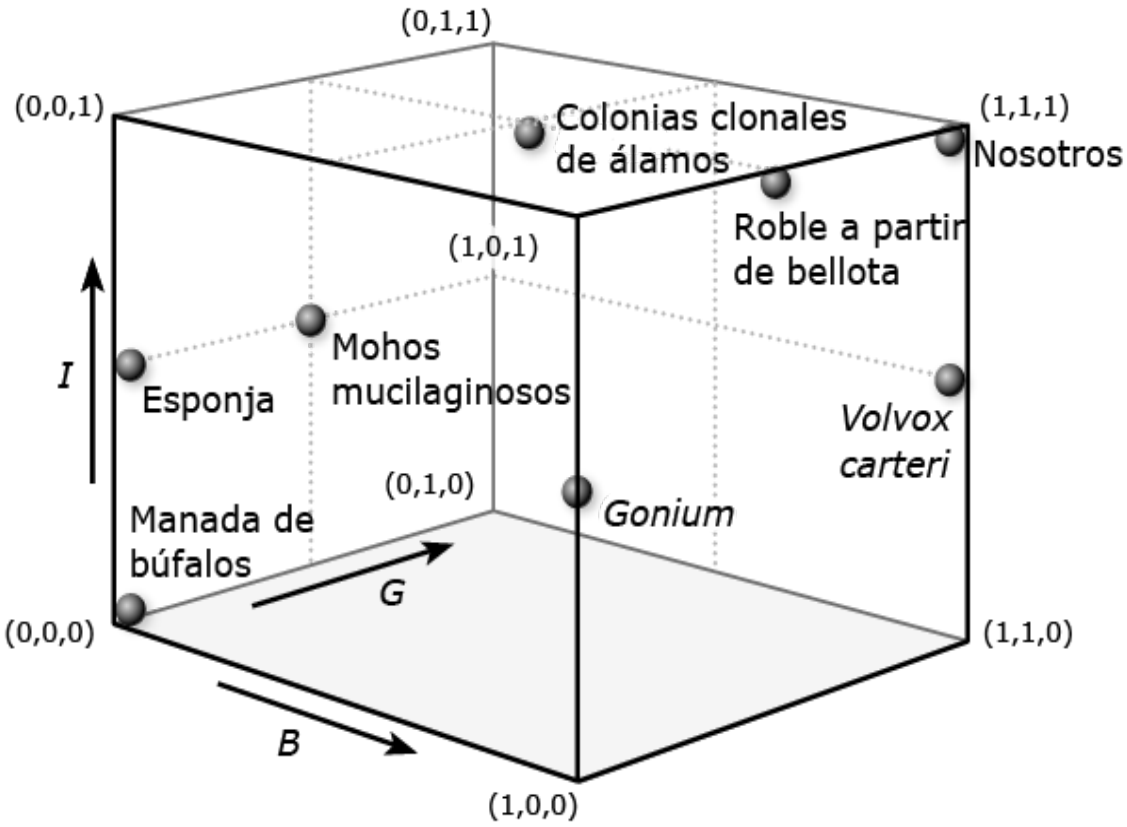


FIGURA 7.2. Espacio darwiniano con otras variables. © Godfrey-Smith

La figura 7.3 es una aplicación de los espacios darwinianos al origen de la vida, tal como hemos comentado durante los tres primeros capítulos. Así, tiene que ser un conjunto de procesos que nos lleve de una situación predarwiniana a otra protodarwiniana y, finalmente, a una darwiniana.

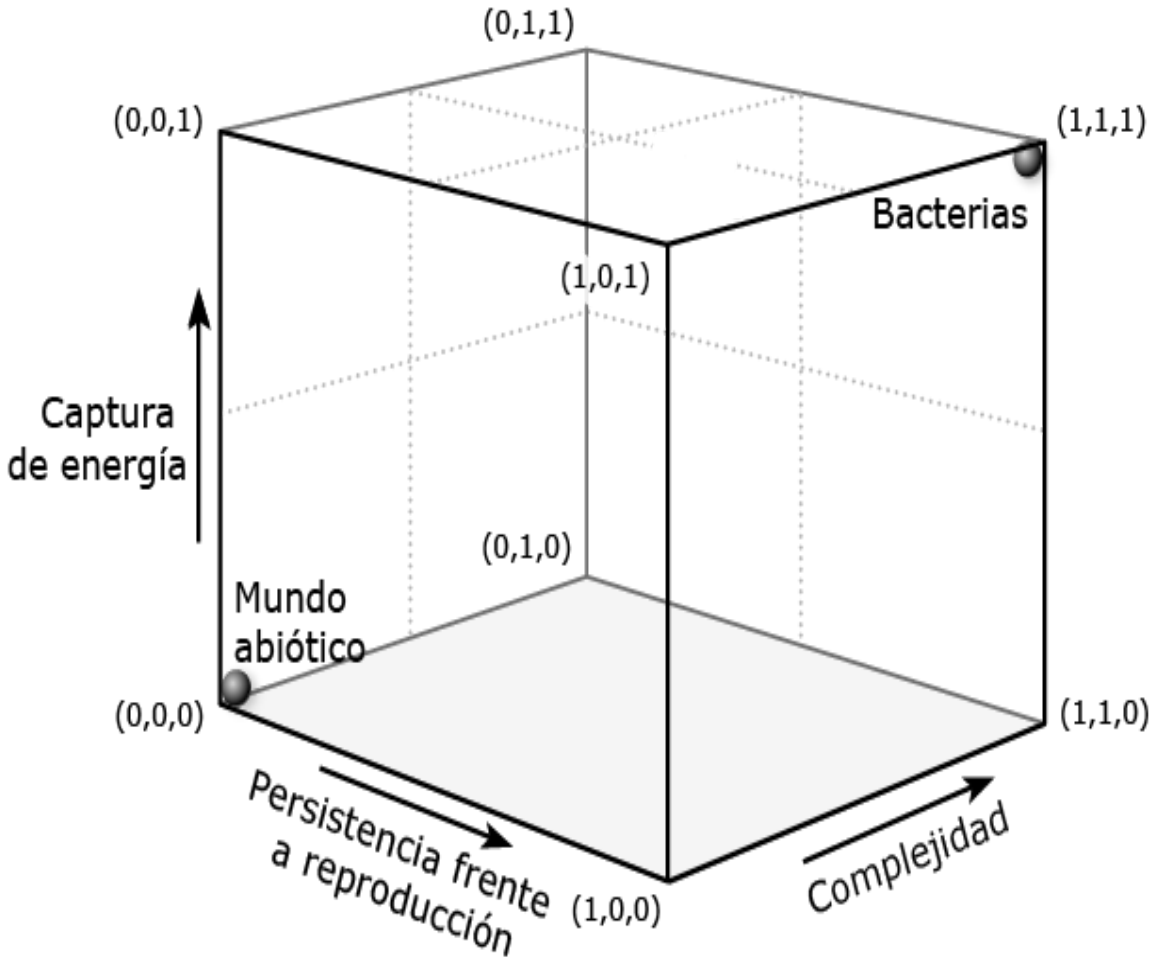


FIGURA 7.3. Espacio darwiniano para el origen de la vida. © Godfrey-Smith

Cuando llegamos a las bacterias, en el vértice superior derecho, tenemos reproducción de pleno derecho, captura de energía y una gran complejidad. La cuestión a responder plantea qué ruta se puede haber seguido hasta llegar ahí. Podríamos representar las trayectorias de los diversos componentes (membrana, metabolismo, mecanismos de replicación, etc.) y ver cuáles se unieron primero y cuáles fueron añadidos posteriores o perfeccionamientos. Se podrían representar otras variables; dos especialmente relevantes quizá serían el tamaño y la eficiencia. Tal como se señaló en el capítulo 2, acaso las primeras cosas o conjuntos que, más o menos, se reprodujeron eran cual inventos del Tebeo, comparativamente enormes y lo bastante estables para que la selección

natural los podara y racionalizara en bacterias compactas y de gran eficiencia. En este momento, resultaría prematuro llenar el espacio con ejemplos; incluso el lugar donde he colocado a las bacterias, el vértice superior derecho, podría ser engañoso si las arqueas (o alguna otra cosa) resultan ser los primeros elementos claramente darwinianos que se reproducían.

EVOLUCIÓN CULTURAL: LA INVERSIÓN DE UN ESPACIO DARWINIANO

Pasemos ahora a considerar una aplicación de los espacios darwinianos al fenómeno de la evolución cultural, dejando para más adelante los detalles con los que explicarla y respaldarla.

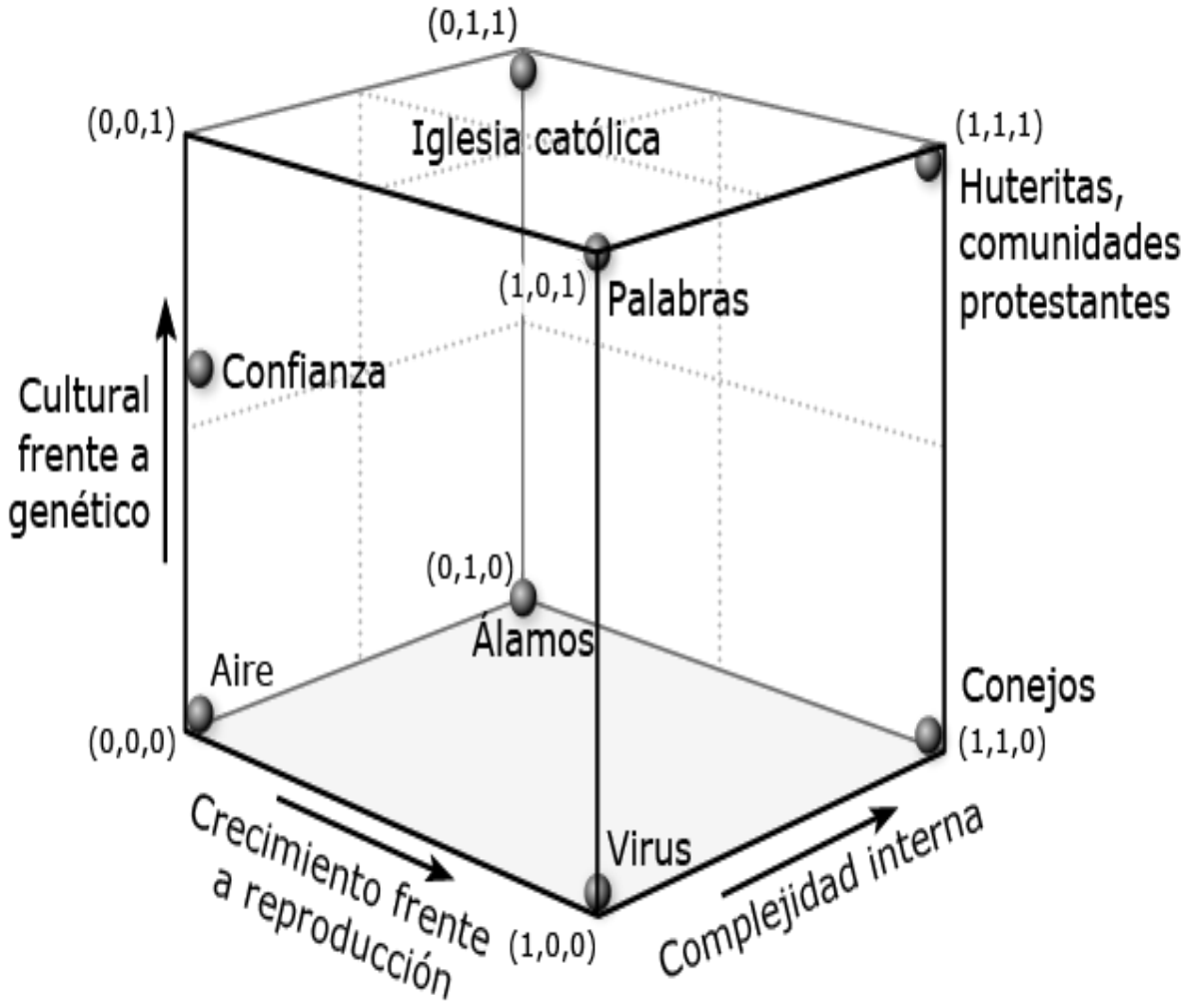


FIGURA 7.4. Espacio darwiniano con religiones. © Godfrey-Smith

En el eje x de la figura 7.4 he representado el crecimiento frente a la reproducción. Las alamedas, normalmente, hacen crecer sus «cuerpos» en lugar de tener «descendencia»; una mezcla de «solo crecimiento» y también algo de reproducción. Las setas son otro ejemplo, pero no sé de animales que puedan duplicarse, cual siameses, en lugar de reproducirse del modo habitual. En el eje z represento el grado de complejidad interna y en el y, la evolución cultural frente a la genética (¿hay fenómenos a medio camino? La respuesta es afirmativa, como veremos). Los mohos mucilaginosos, como los álamos, mezclan la reproducción diferencial con el crecimiento diferencial: creciendo y creciendo en lugar de dividirse en descendientes individuales. En el mundo cultural, la Iglesia católica ha ido creciendo y creciendo (hasta hace poco) pero en la actualidad raras

veces engendra descendientes (aunque produjo una prole espectacular en el siglo XVI). En cambio, los huteritas****** están diseñados para encomendar a sus miembros que funden comunidades hijas cada vez que la comunidad original se vuelve demasiado grande (en Wilson y Sober 1995 se pueden consultar detalles sobre los procesos evolutivos). Las religiones, o las comunidades religiosas, son entes sociales grandes y complejos. Las palabras, en cambio, se parecen más a los virus: simples, sin vida y dependientes de sus anfitriones para la reproducción (no debéis caer en el error habitual de pensar que todos los virus son dañinos: por cada virus perjudicial, hay millones que al parecer son inocuos y algunos pueden ser beneficiosos e incluso esenciales).

La confianza es, mayormente, un fenómeno cultural (sin duda, recibe un empujón de nuestros genes) y, como el aire que respiramos, no es algo demasiado tangible ni cuantificable, no es muy «cosificable», hasta que nos quedamos sin. En gran medida, es un producto de la evolución cultural, del mismo modo que el aire que respiramos es un producto de la evolución genética a lo largo de miles de millones de años. Cuando surgió la vida, esta era anaeróbica (no necesitaba oxígeno) y en la atmósfera casi no había oxígeno, pero cuando apareció la fotosíntesis, los seres vivos empezaron a expulsar oxígeno a la atmósfera, en forma de CO2 y de O2. Esto llevó varios miles de millones de años y parte del O2 de la estratosfera reaccionó para convertirse en O3, ozono, sin el cual una radiación mortífera llegaría a la superficie de la Tierra e imposibilitaría nuestro tipo de vida. La concentración de oxígeno hace 600 millones de años era solamente el 10% de su valor actual, de modo que aunque el cambio es imperceptiblemente lento, a lo largo del tiempo resulta espectacular. Podemos suponer que la atmósfera es un rasgo fijo del entorno selectivo en que se produce la evolución, pero en realidad también evoluciona y es un producto de toda la evolución previa que se ha producido en el planeta. Estos fenómenos (podemos llamarlos la «atmósfera» biológica y cultural en que se produce la evolución) no se reproducen a sí mismos, pero fluctúan localmente y evolucionan de un modo no darwiniano.

La figura 7.5 es otro avance. En este diagrama he invertido los vértices de los diagramas darwinianos anteriores: los fenómenos darwinianos paradigmáticos se hallan en el vértice (0,0,0) y los fenómenos no darwinianos paradigmáticos, casos del diseño inteligente (no del Diseño Inteligente, esto no tiene nada que ver con la religión) están en el vértice superior derecho.

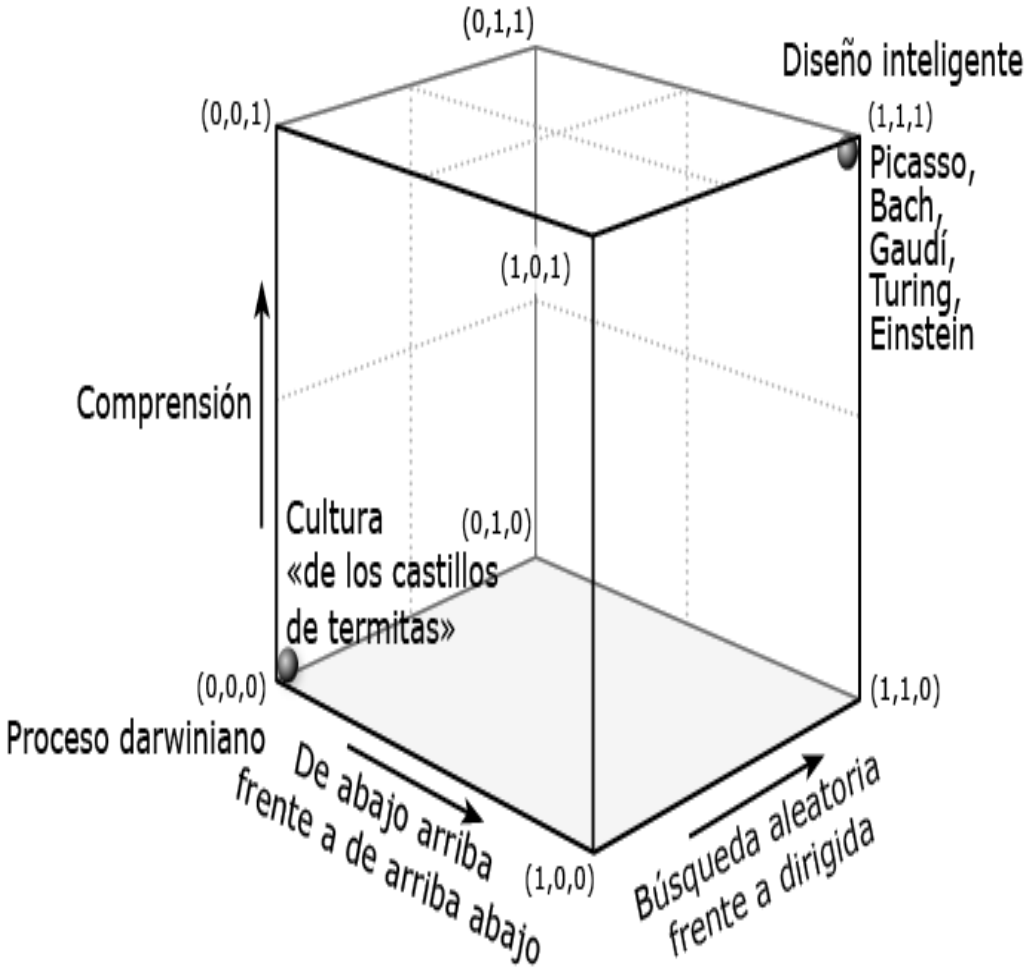


FIGURA 7.5. Espacio darwiniano invertido, con fenómenos darwinianos en (0,0,0) y diseño inteligente en (1,1,1). © Godfrey-Smith

Mi proposición es que la cultura humana se inició como un fenómeno profundamente darwiniano, con competencias sin comprensión que generaban diversas estructuras valiosas de modo parecido a como las termitas construyen sus castillos, y luego, poco a poco, se desdarwinizó, aumentando su comprensión y su capacidad de organización de arriba abajo, aumentando su eficiencia para buscar por el espacio de diseño. En resumen, a medida que la cultura humana evolucionaba, se alimentaba de los frutos de su propia evolución y aumentaba sus poderes de diseño usando la información de maneras cada vez más potentes.

En el vértice superior derecho, el extremo del diseño inteligente, tenemos la situación ideal que, en realidad, nunca se alcanza: el genio divino, el Diseñador Inteligente que, desde su atalaya, otorga dones al resto de mortales. Todos los fenómenos culturales reales se hallan en algún lugar intermedio, con una comprensión imperfecta, una búsqueda imperfecta y gran cantidad de colaboración más o menos mediocre.

He colocado a Picasso en la cumbre no porque crea que fuera más inteligente que los otros genios, sino porque una vez dijo «Je ne cherche pas. Je trouve!» (‘No busco; ¡encuentro!’). Es la típica fanfarronada de un genio, que en pocas sílabas resume el siguiente mensaje: «Yo no tengo que escarbar por todas partes probando y errando. No tengo que deambular fatigosamente para subir la cuesta hasta la cima global del diseño, me basta con dar un brinco cada vez hasta el pico del Everest. Lo comprendo todo; ¡incluso comprendo mi propia comprensión!». Una bobada, sin duda, pero una bobada inspirada; y en el caso de Picasso, también una mentira descarada, pues hizo centenares de esbozos con variaciones sobre un mismo tema, mordisqueando por el espacio de diseño hasta encontrar algo que consideraba una pausa agradable en su viaje (era un gran artista, pero parte de su genio radicaba en su habilidad para firmar, y vender, muchas de sus pruebas, en lugar de descartarlas sin más).

Quedan muchas más cosas por explorar en esta evolución de la evolución cultural y en su papel en la creación de nuestras mentes, pero antes deberíamos echar un vistazo más profundo a cómo empezó. Igual que el origen de la vida, se trata de un problema aún sin solución, y muy difícil de resolver. En algunos aspectos, nuestras mentes son tan diferentes de otras mentes como puedan serlo los seres vivos de los no vivos; hallar tan solo un posible camino que parta de nuestro ancestro común con los chimpancés, y respaldarlo, es una tarea abrumadora. No es que falten hipótesis competidoras de las que aquí plantearemos las que, en mi opinión, son las más prometedoras. ¿Qué fue primero? ¿El lenguaje, la cooperación, la fabricación de herramientas, el dominio del fuego, el lanzamiento de piedras, la búsqueda de comida con enfrentamiento, el intercambio…? No debería sorprendernos si resulta que no hay una única respuesta definitiva sino más bien un proceso coevolutivo con muchos factores que contribuyen y que se nutren mutuamente. En cualquier caso, el hecho es que somos la única especie que, hasta el momento, ha establecido una cultura que se acumula de forma explosiva (otra bomba más de replicación) y cualquier historia especulativa que explique por qué

tenemos cultura (tanto el cómo es que como el para qué) tendrá que explicar también por qué somos los únicos que la tenemos. Parece fuera de toda duda que la cultura ha sido un Buen Truco para nosotros, pero la visión y el vuelo también son unos Buenos Trucos muy evidentes que han evolucionado varias veces en especies diferentes; ¿con qué barreras se han topado las otras estirpes evolutivas que les han impedido desarrollar el Buen Truco de la cultura?

****** Los huteritas son una rama del anabaptismo, similar a los amish y los menonitas. (N. del t.)

8

CEREBROS FORMADOS POR CEREBROS

COMPUTADORAS DE ARRIBA ABAJO Y CEREBROS DE ABAJO ARRIBA

Una bacteria está pertrechada para detectar gradientes que le resultan relevantes, para identificar unas pocas diferencias vitales, para sentirse a gusto en los familiares entornos de su diminuto Umwelt, obteniendo la energía y los materiales que necesita para crecer y reproducirse. Las plantas y otros organismos sésiles organizan las células que los forman en grandes ejércitos de esclavos con tareas concretas a realizar, que los alimentan y los protegen a cambio de sus servicios como supervisores, facilitadores, controladores del crecimiento, etc. sin necesidad de invertir en los sistemas de percepción a distancia que permiten que un organismo pueda ser móvil sin ir chocando con innumerables cosas nocivas. Un sistema de control rápido es una competencia fundamental de cualquier organismo móvil, de ahí que los sistemas nerviosos con un cuartel general sean obligatorios (fijémonos que, aunque las plantas tienen sistemas complejos de transmisión de información que controlan sus respuestas a variaciones en el entorno, se trata de sistemas distribuidos, no centralizados alrededor de algún tipo de núcleo). Los cerebros son centros de control que se ocupan con presteza y de manera adecuada de las oportunidades y los riesgos (los ofrecimientos) de una vida móvil.

Tal como hemos señalado, la selección natural ha diseñado los cerebros para disponer de equipos (o para desarrollarlos con fiabilidad) que puedan extraer la información semántica que necesitan para sus tareas de control.

Un insecto suele nacer con un conjunto maduro de competencias, con sus

«expectativas» ya instaladas gracias a la historia de encuentros de sus antecesores. Un pez ovíparo no tiene tiempo para aprender a nadar cuando es joven, ni tampoco unos padres que le enseñen; ha de nacer con un «instinto natatorio» incorporado. Un ñu recién nacido tiene que arrancar a toda velocidad, con escasas oportunidades para explorar el mundo y descubrir sus maravillas; si no es capaz de seguir a la manada, pronto será comida y carroña para otros. Otros mamíferos, y también los pájaros, pueden permitirse ser altriciales (por oposición a precoces); están diseñados para que los padres los alimenten y los protejan durante una infancia dilatada, en la que van recogiendo información semántica que no tiene que llegarles a través de sus genes ni tiene que aprenderse por un peligroso proceso de ensayo y error sin protección. Aunque son criaturas puramente skinnerianas, pueden probar una selección de entornos de aprendizaje con pocos riesgos, entornos más o menos seguros escogidos por sus padres (de una forma competente pero sin mucha comprensión). De un modo u otro, los cerebros desarrollan competencias, incluyendo las metacompetencias necesarias para adquirir y perfeccionar otras competencias.

Antes de pasar a algunas de las maneras especializadas en que los cerebros están diseñados para extraer información semántica, es momento de abordar el problema de la diferencia tajante que hay entre los cerebros y las computadoras que han inundado nuestro mundo. Muchas tareas de control que hasta hace poco realizaban los cerebros humanos han sido asumidas por computadoras, que se ocupan de ascensores, aviones y refinerías de petróleo. La idea teórica de Turing, puesta en práctica por John von Neumann con la computadora programable en serie, ha crecido exponencialmente en los últimos sesenta años y hoy en día está presente en todos los entornos del planeta, además de haber enviado miles, o quizá millones, de descendientes al espacio; sin duda, el invento más viajado de la historia. Las brillantes idealizaciones de Shannon, Turing, von Neumann, McCulloch y Pitts han conducido a una explosión tan grande en la competencia de manipulación de información que actualmente es habitual suponer que los cerebros no son más que computadoras digitales orgánicas y, además, que pronto las computadoras basadas en el silicio incorporarán una inteligencia artificial que sobrepasará al cerebro humano en todos «los logros de la habilidad creadora» (para citar nuevamente la indignada diatriba de Beverley contra Darwin, acusándole de pensar que la «ignorancia absoluta» podía hacer lo mismo). Creo que la posición de

Beverley queda refutada; la ignorancia absoluta de la evolución por selección natural no solamente puede crear margaritas y peces, sino también seres humanos que, a su vez, tienen la competencia para construir ciudades, teorías, poemas, aviones y computadoras que, a su vez, podrían, en principio, lograr inteligencias artificiales con habilidades creadoras superiores a las de sus creadores humanos.

Pero el tipo de inteligencia artificial que se concibió originalmente para ejecutarse en las maravillosas máquinas de Turing y von Neumann (la GOFAI) seguramente no es la clase adecuada de software que se necesita para una tarea tan trascendental, mientras que el hardware sobre el que se sustenta (la máquina de von Neumann y sus miles de millones de descendientes)[1] puede que no sea la mejor plataforma. Ya comentamos en el capítulo 4 que Turing era el epítome de un diseñador inteligente de arriba abajo, y la computadora que inventó era un utensilio ideal para poner en práctica diseños de arriba abajo de todo tipo. El proyecto de sistema de control del ascensor era un ejercicio de resolución de problemas de arriba abajo: a los programadores se les facilitaba las especificaciones finales de buen principio; de este modo, podían usar su inteligencia para prever las dificultades que podrían surgir, imaginando ciclos y ciclos de trabajo del ascensor y vigilando los ofrecimientos positivos y negativos. ¿Cuándo debe detenerse el ascensor y recoger pasajeros, al subir o al bajar? ¿Qué debe hacer cuando recibe peticiones simultáneas? ¿En qué condiciones puede cambiar de sentido sin hacer vomitar a los ocupantes? Los diseñadores del programa eran criaturas popperianas probando hipótesis en un entorno hipotético, pero también criaturas gregorianas usando un conjunto de recursos de pensamiento para realzar sus poderes, y ningún recurso era más potente que el lenguaje de programación en que decidieran trabajar. La grandeza de un lenguaje de programación es que una vez que el diseño está escrito con claridad en alguno de ellos (como Java, C++ o Python, por ejemplo), se puede confiar en que el programa compilador se ocupará del resto: creará un archivo ejecutable en lenguaje de máquina.[2] La programación informática no es tanto un diseño de arriba abajo como un diseño «de arriba a media altura»; los detalles más farragosos de la parte «de abajo» del diseño (de la sala de máquinas, podríamos decir) se pueden ignorar con tranquilidad, excepto si el programa que estamos escribiendo es precisamente un compilador.

No hay nada que se parezca a esta comodidad para el diseñador en la evolución por selección natural, pero tal como señaló Herbert Simon hace ya muchos años en su brillante librito Las ciencias de lo artificial (1969), los sistemas complejos evolucionables (básicamente, todos los sistemas vivos evolucionables) dependen de que estén organizados jerárquicamente, es decir, formados por partes que tienen una cierta estabilidad, independientemente del sistema más grande del cual forman parte, y estas partes, a su vez, formadas por partes también estables formadas por partes. Una estructura o un proceso solo necesita ser diseñado una única vez, y luego utilizarse y copiarse una y otra vez, no solo entre un organismo y su descendencia, sino también dentro de un organismo a medida que se desarrolla. Richard Dawkins ha señalado que un gen es como una subrutina de caja de herramientas o toolbox en una computadora:

El Mac tiene un repertorio de rutinas toolbox almacenadas en la memoria ROM o en archivos de sistema que siempre se cargan al encender el ordenador. Hay miles de rutinas de este tipo, cada una de las cuales realiza una operación concreta, una y otra vez, de maneras ligeramente distintas, en programas diferentes. Por ejemplo, la rutina llamada «ocultar cursor» hace que el puntero desaparezca de la pantalla hasta que se vuelva a mover el ratón. El gen de «ocultar cursor» se activa, sin que el usuario lo vea, cada vez que el usuario empieza a escribir y el puntero del ratón desaparece. (Dawkins 2004, p. 260)

Esta sistematicidad jerárquica se halla en todos los rincones de la naturaleza; está en el genoma y en los procesos de desarrollo que el genoma controla. Si ejecutamos la subrutina de creación de vértebras unas cuantas veces más, pronto tendremos una serpiente. Las personas que tienen seis dedos (una mutación bastante frecuente) tienen una subrutina de creación de manos que ejecutó su subrutina de creación de dedos una vez más de lo habitual. En cuanto el programa para crear párpados no contiene errores, se puede usar con pequeños cambios en todo el mundo animal. La evolución por selección natural, pues, es un poco como un código fuente, ya que contiene instrucciones relativamente simples que desencadenan una cascada de otras instrucciones para

conseguir una determinada tarea «modular», pero todo está en «lenguaje de máquina» que se ejecuta siempre que se necesita; no tiene que compilarse y no hay ningún lector de este código que necesite ayudas mnemotécnicas para entenderlo (como PLANTALLAMADA, PESOTOTAL, etc. como vimos en el capítulo 4). Es como si el organismo en desarrollo comprendiera las órdenes de sus genes del mismo modo que una máquina de von Neumann comprende sus instrucciones en lenguaje de máquina: las obedece (por decirlo de alguna manera).

COMPETICIÓN Y COLABORACIÓN EN EL CEREBRO

La evolución por selección natural no es una I+D de arriba abajo, como la programación informática, a pesar de su inventiva y abundante uso de módulos. Se trata de una I+D de abajo arriba, es la extraña inversión de Darwin. Además, los cerebros se diferencian de las computadoras digitales en muchos aspectos; los más citados son tres, pero, en mi opinión, no son muy importantes:

Los cerebros son analógicos; las computadoras, digitales. Esto quizá sea cierto, según lo que se considere como analógico; si se quiere decir no binario (que solo usa unos y ceros, encendidos y apagados) entonces seguramente los cerebros son analógicos, pero puede que sean digitales en otros sentidos. Cualquier alfabeto finito de señales con clases de equivalencia es un tipo de digitalización (A, a, a, A, A, a… todas cuentan como ‘a’). EstoeS uNa sEÑalDigITAaliZadA y, como veremos en el próximo capítulo, este tipo de digitalización es un rasgo muy relevante del lenguaje.

Los cerebros procesan en paralelo; las computadoras, en serie. Es decir, los cerebros ejecutan millones de «operaciones» simultáneamente, extendidas por todo el tejido cerebral, mientras que las computadoras ejecutan una única orden simple a la vez, una tras otra, en una cadena

lineal de operaciones que compensa su estrechez con una velocidad fulminante. Hay excepciones; se han construido algunas computadoras para propósitos concretos con una arquitectura en paralelo, pero las computadoras que actualmente están integradas en miles de objetos, desde despertadores y tostadoras a automóviles están concebidas con una arquitectura en serie, son «máquinas de von Neumann». Es muy posible que tengáis centenares de máquinas de von Neumann ocultas en vuestros electrodomésticos y dedicadas a tareas insignificantes que requieren solo una minúscula parte de su potencia (es mucho más barato integrar una computadora completa en un chip fabricado en masa que construir aparatos diseñados específicamente para cada tarea concreta). Es cierto que la arquitectura del cerebro es en paralelo, y a gran escala, con un sistema de visión de una anchura de un millón de canales, más o menos, pero también es cierto que muchas de las actividades más espectaculares del cerebro se producen más o menos en serie, en el llamado flujo de conciencia, en que las ideas, conceptos o pensamientos se presentan no exactamente en fila india, sino a través de una especie de cuello de botella de von Neumann. Es posible simular una máquina en serie virtual con una arquitectura en paralelo (y es lo que hace el cerebro, tal como explico en La conciencia explicada) y se pueden ejecutar computadoras en paralelo en máquinas que operan en serie, con la extensión que uno desee pero a costa de reducir la velocidad de funcionamiento. No es ninguna coincidencia que los cerebros sean máquinas de procesamiento en paralelo a gran escala, teniendo en cuenta la necesidad de velocidad para controlar vidas en constante peligro, pero las «piezas móviles» de las máquinas actuales de von Neumann tienen un tiempo de ciclo básico miles de millones de veces más breve que el tiempo de respuesta de una neurona. En los últimos años, los investigadores en inteligencia artificial han elaborado redes conexionistas que, en líneas generales, imitan las redes neuronales del cerebro y que han demostrado ser muy buenas aprendiendo a diferenciar cosas en imágenes y usando otros tipos de reconocimiento de patrones, tareas en las que nuestros cerebros son verdaderos maestros. Sin duda, demuestran la potencia del procesamiento en paralelo (véase el capítulo 15) pero, de hecho, el procesamiento en paralelo efectuado por estas redes está simulado casi por completo en máquinas convencionales de von Neumann, con su procesamiento en serie. Las máquinas de von Neumann son imitadores universales que pueden simular cualquier otra arquitectura; las rápidas máquinas de von Neumann de hoy, a pesar de sus drásticos cuellos de botella, pueden hacerse pasar por redes neuronales de procesamiento en

paralelo con una velocidad tal que siguen sin problemas el ritmo de los cerebros de procesamiento en paralelo que las han diseñado (e incluso los superan).

Los cerebros se basan en el carbono (proteínas, etc.); las computadoras, en el silicio. Por el momento, esto es cierto, pero se han logrado avances espectaculares en nanotecnología para construir computadoras a partir de proteínas y, por otro lado, las redes de proteínas del interior de las células parecen hacer tareas computacionales (véase Bray 2009 para una brillante exposición). Nadie ha demostrado todavía que la diferencia en la composición química subyacente tenga que dar ninguna ventaja al carbono.

¿Qué tal esta otra diferencia?

Los cerebros están vivos; las computadoras, no.

Algunos pueden objetar (y es lo que yo mismo solía hacer) que un corazón artificial no está vivo, pero cumple su función sin problema. Una rótula o una cadera sintética no tienen ninguna necesidad de estar vivas. Se puede sustituir o empalmar el nervio auditivo con un alambre, adecuadamente conectado en ambos extremos… ¿por qué no el resto del cerebro? ¿Acaso hay alguna parte superespecial del cerebro que tiene que estar viva para que funcione el conjunto? La suposición habitual de trabajo de la inteligencia artificial siempre ha sido que cualquier órgano vivo no es más que un trozo muy sofisticado de maquinaria basada en el carbono, que se puede sustituir, pieza a pieza o todo de vez, por un sustituto no vivo con el mismo perfil de entradas y salidas (es decir, que hace las mismas cosas cuando se somete a las mismas entradas durante el mismo tiempo), sin perjuicio alguno. Si el cerebro fuera un órgano para secretar bilis o para purificar la sangre, los aspectos químicos y físicos de las partes implicadas serían muy relevantes y quizá podría llegar a resultar imposible hallar sustitutos adecuados de los materiales empleados, pero el cerebro es un órgano de procesamiento de información y la información no depende del medio (una advertencia o

una declaración de amor se pueden «crear de la nada», siempre y cuando el receptor tenga la capacidad para captarlos).

Hay un rasgo de los seres vivos que puede ser de gran relevancia en nuestra búsqueda para comprender cómo funcionan las mentes, un rasgo sobre el que recientemente ha echado luz Terrence Deacon, en su arduo e importante libro Naturaleza incompleta: cómo la mente emergió de la materia (2012). Según Deacon, los mismos avances que he descrito como simplificaciones geniales han llevado la teoría hacia zonas inapropiadas del espacio de diseño durante más de medio siglo. La jugada maestra de Shannon fue abstraer el concepto de información y alejarlo de la termodinámica y de la idea de energía (y de materia, como hemos remarcado); la información es información, da igual qué es lo que esté implicado, si electrones, fotones, señales de humo, regiones magnéticas u hoyos microscópicos en discos de plástico. Se necesita energía para transmitir y transformar información (no se trata de magia, al fin y al cabo), pero podemos aislar nuestra comprensión del procesamiento de información de las eventuales consideraciones energéticas implicadas. Norbert Wiener creó el campo de la cibernética, y también acuñó el término, que tomó del verbo griego que significa ‘pilotar’ o ‘dirigir’ (κυβερνᾶν, kybernan), del cual también deriva el verbo gobernar. Wiener señaló que mientras que un controlador (de un barco, una ciudad, un cuerpo o una refinería de petróleo) necesita usar energía en algún punto (para girar el timón, para emitir un bando, para subir la temperatura), la energía necesaria para hacer funcionar el sistema de control es discrecional: la que quieras y, en cualquier caso, no mucha. Esta separación entre la computación y los aspectos dinámicos del control físico es lo que hoy en día permite controlar cualquier cosa con tu teléfono inteligente, que solo necesita una batería (más transductores y accionadores de potencia suficiente para hacer el trabajo duro). Cuando usas tu teléfono para enviar una señal de radio y así abrir la pesada puerta de tu garaje, la batería consume poquísima energía; es el motor eléctrico que esa señal controla el que utiliza una cantidad considerable de electricidad para efectuar su trabajo. Deacon admite que esto ha permitido la fructificación de todo el diseño inteligente de alta tecnología que hemos explorado, pero insiste en que al separar el procesamiento de información de la termodinámica, nuestras teorías se limitan a sistemas básicamente parasitarios, artefactos que dependen de un usuario para obtener energía, para su mantenimiento estructural, para su interpretación y para su raison d’être. Los seres vivos, en cambio, son

autónomos y, además, están formados por seres vivos (células) que, en cierto modo, también son autónomos.

Los investigadores en inteligencia artificial podrían responder que están posponiendo la consideración de fenómenos como la captura de energía, la reproducción, la autorreparación y la autorrevisión abierta a la espera de cosechar los beneficios de esta simplificación, clarificando primero los fenómenos puramente informacionales del aprendizaje y la autoorientación (una especie de autonomía, por muy incompleta que sea). A nadie le pasaría por la cabeza complicar el diseño de una computadora para jugar al ajedrez exigiéndole que obtuviera su energía a partir de bocadillos y refrescos y obligándole a supervisar no solo el tiempo, sino también la reducción de sus reservas de energía. Los jugadores de ajedrez humanos tienen que controlar sus punzadas de hambre y emociones como la humillación, el miedo o el aburrimiento, pero las computadoras pueden pasar de puntillas con elegancia por todo esto, ¿no es cierto? En efecto, afirma Deacon, pero con un enorme coste: al quitar todas estas preocupaciones, los diseñadores de sistemas crean arquitecturas frágiles (no se pueden reparar a sí mismas, por ejemplo), vulnerables (limitadas al conjunto de contingencias que sus diseñadores hayan previsto) y totalmente dependientes de sus controladores.[3]

Deacon insiste en que esta diferencia no es baladí. ¿Es así? En algunos aspectos, creo que tiene razón. En el momento álgido de la GOFAI, en la década de 1970, observé que los programas de inteligencia artificial solían ser aspirantes a genios, despersonificados y «postrados en cama», que solo se podían comunicar leyendo y escribiendo mensajes a máquina (los trabajos sobre visión por computadora solían hacerse con una única cámara fija o cargando imágenes fijas en el sistema, del mismo modo que cargamos fotografías en nuestro portátil, un sistema de visión sin ojos). Un robot móvil corpóreo con «órganos» de los sentidos para ubicarse a sí mismo en el entorno consideraría que algunos problemas son más fáciles y otros más difíciles. En 1978, escribí una pequeña observación: «¿Y por qué no la iguana completa?» en la que sostenía que deberíamos dejarnos de microcompetencias humanas limitadas (responder preguntas sobre el béisbol, jugar a las damas) y, en su lugar, intentar crear modelos de la arquitectura computacional de un animal robótico completo, autoprotector y con capacidad de capturar energía, por muy rudimentario que fuese (hay que simplificar por algún lugar… la inteligencia artificial no es pan comido); el animal podría ser ficticio,

si con eso se facilitaba la tarea (una iguana marciana de tres patas, por decir algo).

Algunos expertos en robótica aceptaron el reto. Por ejemplo, Owen Holland inició el proyecto SlugBot, para intentar construir un robot que recogiera babosas de los campos de cereales y las digiriese para producir la electricidad necesaria para alimentar sus chips informáticos, un proyecto que ha engendrado otros ingeniosos diseños. Aquí tenemos captura de energía, pero ¿qué hay de la autorreparación y el automantenimiento? Se ha trabajado durante décadas en estos y otros requisitos de la vida, y se han creado animales artificiales (animats) y nanorrobots, que ejemplifican diversas versiones muy y muy simplificadas de los procesos básicos de la vida, pero nadie ha llevado la idea de la «iguana completa» tan lejos como Deacon afirma que se debería haber hecho: en sus laboratorios, estas criaturas artificiales siguen dependiendo de chips con CPU (el hardware de von Neumann) aunque simulen arquitecturas de procesamiento en paralelo. Según Deacon, es importante que un cerebro esté formado por células que sean, por ellas mismas, pequeños agentes autónomos con sus objetivos, el más importante de los cuales es sobrevivir, objetivo que a su vez engendra otros, como encontrar trabajo y encontrar aliados. La insistencia de Deacon en formar cerebros (o suplentes de cerebros) a partir de neuronas vivas puede parecer, a primera vista, una concesión al romanticismo, un chovinismo proteínico, en realidad, un paso fatal que nos aleja del vitalismo, pero sus argumentos son prácticos y convincentes.

Podemos aportar algo de luz a la idea de Deacon gracias a un fenómeno: la considerable plasticidad del cerebro. Si se daña una zona del cerebro, a menudo (aunque no siempre), una zona adyacente puede encargarse de las tareas del tejido dañado, con rapidez y elegancia. Si una zona está infrautilizada, las vecinas no tardan en reclutar células de esa zona para que les ayuden en sus propias tareas. En general, las neuronas que se destruyen no son reemplazadas del modo en que sí lo son las células de la piel, de los huesos o de la sangre, entre otras. La neurorregeneración es todavía un sueño de los bioingenieros, no un rasgo habitual de los sistemas nerviosos, de modo que la plasticidad que se observa en tantos experimentos tiene que deberse a una reasignación de neuronas a nuevas tareas o a una carga adicional de trabajo. ¿Hay algún director de recursos humanos que determina estos nuevos puestos de trabajo y que envía órdenes desde lo alto hacia las neuronas subordinadas que trabajan en el

taller? El informático Eric Baum, en What is Thought? (2004), denomina a este tipo de organización, control del politburó y señala que esta forma de lograr tareas complejas de arriba abajo no es cómo funciona el cerebro. Los economistas han mostrado que las economías planificadas no funcionan tan bien como las economías de mercado y que las arquitecturas con una planificación centralizada (de arriba abajo) son organizadores y cerebros poco eficaces por las mismas razones.

El experto en robótica Rodney Brooks apunta (correspondencia personal) que el hardware de las computadoras digitales existentes depende de forma crítica de millones (o miles de millones) de elementos idénticos, clones perfectos casi hasta la escala atómica, de modo que siempre responderán como se supone que tienen que hacerlo: ¡como robots! Los ingenieros han conseguido crear la tecnología para imprimir circuitos informáticos microscópicos, con millones de elementos biestables cada uno de los cuales puede almacenar con fiabilidad un cero o un uno hasta que se le ordene (desde las alturas) cambiar el bit. Estas son las partes móviles finales de una computadora, sin ningún tipo de individualidad ni idiosincrasia. En cambio, las neuronas son todas diferentes; hay muchos tipos estructurales claramente definidos (piramidales, fusiformes, poliédricas, etc.), pero incluso dentro de un mismo tipo no hay dos neuronas idénticas. ¿Cómo se puede organizar una población tan diversa y generar cosas? Sin duda, no gracias a jerarquías burocratizadas, sino mediante una formación y coalición de abajo arriba, con gran cantidad de competición.

NEURONAS, MULAS Y TERMITAS

Las neuronas, como todas las demás células que forman nuestro cuerpo, son descendientes domesticados de las eucariotas unicelulares e independientes que medraban y se las apañaban por su cuenta en un mundo cruel de organismos unicelulares. Es bien sabido que François Jacob afirmó que el sueño de cualquier célula es convertirse en dos células, pero lo cierto es que las neuronas casi nunca tienen descendencia. Igual que las mulas, las neuronas tienen padres (en realidad, una madre, una abuela y así sucesivamente), pero son estériles,

de modo que su bien supremo es mantenerse con vida en su nicho desdarwinizado. Tienen que trabajar para capturar la energía que necesitan y, si no tienen trabajo a jornada completa, se pondrán a trabajar en cualquier otra cosa que tengan a mano.

Últimamente, en diversos rincones de las ciencias cognitivas han fructificado versiones de esta idea y como me han convencido de abandonar una postura que había defendido durante mucho tiempo, me gustaría subrayar mi conversión. He cambiado de idea acerca de cómo tratar la tentación del homúnculo: las ansias casi irresistibles de colocar a un «hombrecillo en el cerebro» para que ejerza de Jefe Supremo, de Significador Central, de Gozador de placeres y Sufridor de daños. En Brainstorms (1978) describí y defendí la estrategia clásica de la GOFAI que posteriormente se conoció, gracias a Lycan (1987), como funcionalismo homuncular: sustituir al hombrecillo por un comité.

El programador de IA empieza con un problema caracterizado de modo intencional y, así, antropomorfiza la computadora con naturalidad: si resuelve el problema, dirá que ha diseñado una computadora que puede (por ejemplo) comprender preguntas en inglés. La primera escala de diseño, y la más amplia, divide la computadora en subsistemas, a cada uno de los cuales se le asignan tareas caracterizadas de modo intencional; el programador elabora un diagrama de flujo con evaluadores, memorizadores, discriminadores, supervisores, etc. Son homúnculos con creces […] A su vez, cada homúnculo se analiza en homúnculos más pequeños y, lo que es más relevante, en homúnculos menos inteligentes. Cuando se llega a una escala en que los homúnculos no son más que meros dispositivos para sumar y restar, cuando la única inteligencia que necesitan es para escoger el mayor de dos números cuando se les ordena hacerlo, han quedado reducidos a funcionarios «que pueden ser sustituidos por una máquina». (Dennett 1978, p. 80)

Sigo pensando que esta argumentación va por el buen camino, pero ahora rechazo y me arrepiento de las connotaciones que pueden tener dos de los términos que utilicé: comité y máquina. La burocracia cooperativa que sugiere el primero, con sus claras relaciones de información (una imagen potenciada por los sensatos diagramas de flujo de los modelos de

la ciencia cognitiva clásica), capturaba el sueño de una GOFAI de arriba abajo, pero también sugería una especie de eficiencia muy antibiológica. La extraña inversión del razonamiento de Turing sigue en pie: al final, en nuestra cascada de descomposiciones llegamos a elementos cuyas tareas son tan rígidas y rutinarias que «pueden sustituirse por una máquina», igual que un diligente computador humano de Turing. Las partes móviles más simples en el interior de las neuronas (proteínas motrices, microtúbulos, etc.) son, en efecto, autómatas sin propósito, como las escobas marchando en formación de El aprendiz de brujo, pero las neuronas mismas, miles de millones de ellas, tienen papeles mucho más innovadores y singulares que los oficinistas menesterosos y obedientes con los que las comparaba. Esto tiene importantes implicaciones para la arquitectura computacional del cerebro.

Tecumseh Fitch (2008) acuñó el término nanointencionalidad para describir la voluntad que muestran las neuronas; Sebastian Seung, por su parte, en su conferencia inaugural del congreso de la Sociedad para la Neurociencia de 2010 trató el tema de las «neuronas egoístas» y ya antes (2003) había escrito acerca de «sinapsis hedonistas». ¿Qué puede querer una neurona? Pues la energía y las materias primas que necesita para prosperar (igual que sus ancestros, las eucariotas unicelulares, y otros parientes lejanos, como las bacterias y las arqueas). Las neuronas no son muy diferentes a los robots; está claro que no son conscientes de ningún modo mínimamente sofisticado (recordad que son células eucariotas, semejantes a las células de las levaduras o a los hongos; si las neuronas individuales tienen conciencia, también la tendrán los hongos causantes de la tiña del pie). Pero las neuronas, como las levaduras y los hongos, son agentes muy competentes en la lucha a vida o muerte, y no en el entorno que hay entre los dedos de los pies, sino en el que hay entre las orejas, donde la victoria se la llevan las células que pueden formar redes con más eficacia y así contribuir a los cambios más relevantes que se dan a una escala en que se pueden identificar propósitos y necesidades humanas.

Turing, Shannon y von Neumann abordaron un proyecto de ingeniería novedoso y de extraordinaria dificultad: diseñar inteligentemente computadoras en las que las necesidades y la eficacia de las tareas quedaran separadas casi por completo. A la escala del hardware, la energía eléctrica se reparte equitativamente y con abundancia; ningún circuito corre el peligro de morirse de hambre. A la escala del software,

un organizador benévolo distribuye los ciclos de la máquina a los procesos en función de su prioridad, y aunque puede haber algún mecanismo de competición para determinar qué procesos tienen la máxima prioridad, en general se trata de una fila ordenada, no de un «sálvese quien pueda» (una indefinida apreciación de este hecho es lo que quizá lleva a muchas personas a considerar que una computadora nunca puede «preocuparse» por nada, no porque esté hecha de materiales inadecuados —¿por qué debería ser el silicio menos apropiado que las moléculas orgánicas?—, sino porque su economía interna no incorpora riesgos ni oportunidades y por ello sus diversas partes no tienen que preocuparse de nada).

La arquitectura jerárquica de arriba abajo del software informático, respaldada por un sistema operativo con abundancia de organizadores, controladores y diversos guardias de tráfico, es un buen ejemplo de la máxima de Marx «a cada cual según sus necesidades, de cada cual según sus capacidades». Ningún circuito sumador ni biestable tiene que preocuparse de dónde recibirá la energía eléctrica que necesita para ejecutar sus tareas y no hay lugar para la «promoción». En cambio, una neurona, siempre está sedienta de trabajo; alarga ramificaciones dendríticas para explorar, buscando conectarse en red con sus vecinas de alguna manera que sea beneficiosa para ella. Así, las neuronas son capaces de autoorganizarse en equipos que pueden encargarse de procesos importantes de manipulación de información, a punto para recibir nuevas tareas que dominarán con una pizca de ensayos de prueba y error. Esto se parece bastante a los miles de mujeres sin trabajo que se abalanzaron a las puertas de Oak Ridge, listas para aceptar tareas que no tenían necesidad alguna de comprender; pero hay una diferencia relevante: no hay ningún general Leslie Groves que diseñe el cerebro, este tiene que diseñarse de abajo arriba.

Ha llegado el momento de evaluar lo que implican todas estas consideraciones. El diseño inteligente de arriba abajo de las computadoras clásicas ha creado unos portentos hipercompetentes pero profundamente abiológicos, y no porque estén hechos del material inapropiado, sino porque este material está organizado con una estructura inadecuada: es el tipo de burocracia planificada que puede funcionar como una máquina bien engrasada pero que depende de una reglamentación estricta de las partes móviles, lo que anula la exploración y la improvisación a todas las escalas. Tenemos que subrayar, una vez

más, que, a priori, esta conclusión deja la puerta abierta a cerebros de silicio (o, en general, de Turing, independientemente del material); las neuronas virtuales o simuladas son posibles en principio, pero muy caras desde el punto de vista computacional; se tendría que modelizar cada neurona con todas sus particularidades, costumbres, preferencias y debilidades. Para mirarlo con un poco de perspectiva, imaginemos que unos científicos intergalácticos descubrieron nuestro planeta y empezaron a estudiar a distancia el «comportamiento» de nuestras ciudades y pueblos, de nuestras autopistas, ferrocarriles y sistemas de comunicación, como si fueran algún tipo de organismo. Grandes amantes de los modelos informáticos, deciden hacer una simulación de Nueva York, con todas sus actividades: Nuevayorkbot. Los escépticos les advierten: «Intentadlo si queréis, pero daos cuenta de que tendréis que modelar los millones de ciudadanos (esas cosillas fofas que se mueven en su interior) con mucho detalle si queréis tener alguna esperanza de lograr un buen modelo predictivo; no hay dos iguales y tienen curiosidad e iniciativa».

Los diseños inteligentes de arriba abajo dependen de la previsión, una cosa de la que la evolución carece totalmente. Los diseños generados por la selección natural siempre son, de algún modo, retrospectivos: «esto es lo que ha funcionado en el pasado». Si también funcionará en el futuro depende de la persistencia de las regularidades en que se basa cada diseño. El diseño de la polilla se puede basar en la regularidad del Sol y de la Luna como únicas fuentes de luz de su entorno, de modo que la costumbre «vuela a un ángulo constante de n grados respecto a la luz» era una buena instrucción a programar en las polillas, hasta que aparecieron las velas y las bombillas eléctricas. He aquí un ejemplo de que demasiada regularidad en el entorno puede convertirse en una trampa, ya que encadena a toda una estirpe con una determinada técnica o propensión que puede resultar letal si cambian las condiciones. Tal como vimos en el capítulo 5, los entornos selectivos variables, a causa de su impredecibilidad, favorecen la selección de diseños incompletos, a los que se añaden mecanismos para ajustarlos y, así, adecuarse a las circunstancias: plasticidad explotable o, para usar el término habitual, aprendizaje. Los sistemas al estilo de la GOFAI, con una estructura de arriba abajo, pueden equiparse con parámetros ajustables y con capacidades de aprendizaje, sin duda, pero la jerarquía de control tiende a comprometerlos a un intervalo fijo de plasticidad basado en una estimación del peor de los casos posibles que puede producirse.

La evolución por selección natural, sin previsión, siempre con la mirada hacia atrás, no es diseño inteligente, pero sí que es una potente I+D, lo bastante perceptiva para establecer la división general del trabajo del cerebro en las recetas genéticas acumuladas y perfeccionadas durante miles de millones de años de vida móvil. Y, como ya hemos visto, las mejoras y variaciones en los genes pueden establecer los parámetros suficientes para garantizar que los comportamientos instintivos complejos (como la construcción de nidos) sobrevivan una generación tras otra. Fijar todos los detalles a través de instrucciones genéticas sería imposible; se necesitarían muchísimos más bits de información de Shannon de los que se pueden transmitir en un genoma, incluso en uno con tres mil millones de nucleótidos, como el nuestro. Este es el mensaje indiscutible de la biología evolutiva del desarrollo (o la evo-devo, como se la conoce informalmente), que remarca que la producción de la siguiente generación de un organismo no es un asunto directo consistente en seguir un plano o una receta descrita en los genes; es un trabajo de construcción que debe basarse en la I+D local que depende de la actividad de agentes locales (miopes) que efectúan ensayos más o menos aleatorios al interactuar con su entorno durante el desarrollo. Las neuronas, por ejemplo, son como agentes skinnerianos, que se arriesgan y aprovechan al máximo las oportunidades que se les presentan, explotando su propia plasticidad para mejorar los procesos por los que reciben recompensas o refuerzos. Sin embargo, sus esfuerzos reciben la ayuda de algunos puntos de referencia útiles dispuestos por los genes, que les indican ciertos caminos. Los cerebros se parecen más a colonias de termitas que a empresas o ejércitos diseñados inteligentemente.

¿CÓMO RECOGE EL CEREBRO LOS OFRECIMIENTOS?

Un tema recurrente del libro ha sido que todos los organismos, desde las bacterias hasta nosotros, están diseñados para tratar con un conjunto de ofrecimientos, las «cosas» que les importan (en un sentido amplio de cosa), y que este catálogo, el Umwelt del organismo, se rellena mediante dos procesos de I+D: la evolución por selección natural y algún tipo de aprendizaje individual. Gibson no dijo nada sobre cómo los organismos

recogen la información necesaria para captar, identificar y supervisar sus ofrecimientos; yo también he estado aplazando el problema, hasta ahora.

Lo que hemos descubierto sobre el predicamento del organismo es esto: está flotando en un mar de diferencias, unas pocas de las cuales pueden marcar la diferencia para él. Desciende de una larga estirpe de vencedores y llega pertrechado con mecanismos y preferencias para filtrar y ajustar las diferencias más relevantes, separando la información semántica del ruido. En otras palabras, está preparado para salir adelante en algunos aspectos; tiene expectativas incorporadas que han prestado un buen servicio a sus antecesores, pero que puede ser necesario revisar en cualquier momento. Decir que tiene estas expectativas es lo mismo que decir que viene equipado con respuestas adecuadas parcialmente prediseñadas, a punto para usarlas. No tiene que gastar su valioso tiempo descifrando, a partir de principios básicos, qué hacer con A, con B o con C; son problemas familiares ya resueltos que asocian un dato inicial con una respuesta, una percepción con una acción. Estas respuestas a los estímulos de sus órganos de los sentidos pueden ser comportamientos externos: un pezón ofrece dar de mamar, las extremidades ofrecen movimiento, un enfrentamiento doloroso ofrece retirarse; pero también pueden ser respuestas ocultas, internas, que convierten a los ejércitos neuronales en equipos eficaces para enfrentarse a futuras tareas.

¿Cómo se produce esta formación, este entrenamiento? En este punto vale la pena recuperar una distinción algo desacreditada de los primeros días de la ciencia cognitiva: modelos de competencia frente a modelos de ejecución. Un modelo de competencia (como la gramática de una lengua) dice cómo debería funcionar el sistema (da unas especificaciones, como en el ejemplo del ascensor del capítulo 4) y deja los detalles de cómo conseguirlo como un problema de diseño con muchas posibles soluciones, con muchos modelos de ejecución. Un modelo de ejecución para aplicar una gramática, para crear un hablante de inglés correcto, por ejemplo, sería un trabajo de la neurolingüística para un futuro lejano. En la primera época de la ciencia cognitiva, los lingüistas teóricos discutían sobre gramática sin preocuparse de cómo podía ser posible que un cerebro «siguiera las normas»; les bastaba con lograr primero las normas correctas.

Mientras tanto, los psicolingüistas ideaban ingeniosos experimentos que mostraban ciertos patrones en los errores gramaticales cometidos por los

niños, fuentes de interferencia, fallos de apreciación gramatical y cosas parecidas. Cuando los teóricos no pudieron explicar estos patrones, se justificaron diciendo que era demasiado pronto para intentar comprender los desagradables detalles de la ejecución. La gente comete errores, a la gente le falla la memoria, la gente puede sacar conclusiones apresuradas, pero la competencia básica sigue estando a buen recaudo en sus gramáticas, por muy descuidada que sea su ejecución en tal o cual ocasión. Los modelos de ejecución llegarían más adelante.

Esta división no resultó siempre útil y creó una brecha en la ciencia cognitiva que aún provoca tensiones y problemas de comunicación. Los lingüistas estaban en lo cierto al insistir que hasta tener una idea bastante clara de lo que un cerebro puede hacer y debería hacer para percibir y expresar lenguaje, los esfuerzos a aplicar ingeniería inversa a la maquinaria de manipulación lingüística del cerebro serán propensos a ir por caminos equivocados, abordando problemas mal planteados. Pero parte de las reglas de cualquier estudio de ingeniería inversa tiene que ser determinar las limitaciones y la solidez de la maquinaria disponible, y al aplazar las preguntas sobre el cerebro y subestimar la importancia de las reveladoras ejecuciones descubiertas por los psicolingüistas, los teóricos también se lanzaron a buscar agujas en pajares.

Ahora que ya estamos avisados, podemos abordar la idea que actualmente está propagándose por la ciencia cognitiva como una respuesta muy prometedora a la pregunta de cómo el cerebro recoge y usa la información semántica disponible: la codificación predictiva jerárquica bayesiana (en Hinton 2007 y Clark 2013 podéis encontrar excelentes descripciones, así como los comentarios de Clark y Hohway 2013). La idea básica es exquisita. El reverendo Thomas Bayes (1701-1761) ideó un método para calcular probabilidades a partir de las expectativas previas propias. Cada problema se formula de esta manera: teniendo en cuenta que tus expectativas basadas en la experiencia previa (y aquí podemos incluir la experiencia heredada de tus ancestros) son tal y cual (expresadas como probabilidades de cada alternativa), ¿qué efecto deberían tener los siguientes datos nuevos en tus expectativas futuras? ¿qué ajustes de tus propias probabilidades serían racionales? La estadística bayesiana es una disciplina normativa, que supuestamente indica la forma correcta de pensar acerca de las probabilidades.[4] Así pues, es un buen candidato para un modelo de competencia del cerebro,

ya que funciona como un órgano que genera expectativas, creando nuevos ofrecimientos sobre la marcha.

Pensemos en la tarea de identificar símbolos escritos a mano (letras y números). No es casualidad que esta tarea suela usarse en Internet como una prueba para distinguir entre seres humanos reales y bots programados para invadir sitios web: la percepción de la escritura, como la del habla, ha demostrado ser una tarea fácil para los humanos pero un reto descomunal para las computadoras. Actualmente, se han elaborado modelos informáticos funcionales que identifican bastante bien números manuscritos (de hecho, garabateados). Estos modelos se basan en diversas capas en cascada: las superiores hacen predicciones bayesianas sobre lo que la siguiente capa inferior del sistema «verá» a continuación; cuando las predicciones se demuestran falsas, generan señales de error como respuesta, lo que conduce a una revisión bayesiana que se incorpora nuevamente en el sistema, una y otra vez, hasta que este se estabiliza en una determinada identificación (Hinton 2007). La práctica lleva a la perfección y, a medida que pasa el tiempo, estos sistemas se hacen cada vez mejores en su tarea, tal como hacemos nosotros mismos (aunque mejor; véase el capítulo 15).

La codificación predictiva jerárquica bayesiana es un método para generar ofrecimientos en grandes cantidades: esperamos que los objetos sólidos tengan una parte posterior que podremos ver si les damos la vuelta; esperamos que las puertas se abran; esperamos que las escaleras permitan subir y que las tazas puedan contener líquidos. Todas estas previsiones surgen de una red que no se limita a esperar pasivamente para recibir información, sino que está siempre haciendo suposiciones probabilistas acerca de lo que va a recibir como datos procedentes del nivel que tiene por debajo, teniendo en cuenta lo que ya ha recibido y los errores en sus suposiciones anteriores y considerando todo esto la fuente básica de nueva información, como una forma de ajustar sus expectativas previas para la siguiente ronda de suposiciones.

De todas las interesantes propiedades de estas aplicaciones del pensamiento bayesiano al problema de cómo aprenden los cerebros, una muy atractiva es que ofrece una explicación simple y natural de un hecho neuroanatómico desconcertante: en las vías ópticas, por ejemplo, hay más vías descendentes que ascendentes, más señales de salida que de entrada. Desde este punto de vista, la estrategia del cerebro es crear

continuamente «modelos forward» o anticipaciones probabilísticas y usar las señales entrantes para podarlas y mejorar su precisión, si es necesario. Cuando el organismo está en racha, en territorio familiar, las correcciones que llegan se reducen a unas pocas gotas y las suposiciones del cerebro, sin nada que se oponga a ellas, le dan una buena ventaja en lo siguiente que tenga que hacer.

Estos modelos bayesianos proceden de los modelos de «análisis por síntesis» de la ciencia cognitiva primitiva, en los cuales la curiosidad de arriba abajo («¿es un ciervo? ¿es una musaraña?») guía la formación de hipótesis que se deberán comprobar con los datos entrantes (tu cerebro analiza los datos haciendo suposiciones, sintetizando una versión de lo que estás buscando y cotejándolo con los datos). En La conciencia explicada (Dennett 1991, p. 22 y ss.) presenté un modelo especulativo de los sueños y las alucinaciones basado en el análisis por síntesis, en el que argumentaba que la creación de contenido en estos fenómenos podía deberse a nada más que «un proceso aleatorio o arbitrario de confirmación y refutación» (p. 24). Esta idea ha recibido un respaldo espectacular y una mejora considerable gracias a trabajos recientes en Google Research (p. ej. Mordvintsev, Olah y Tyka 2015). Ahora puedo simplificar mi descripción: en una red bayesiana el silencio cuenta como confirmación. Sea cual sea la suposición de alto nivel, cuenta como realidad por defecto en ausencia de refutación.

Desde nuestro punto de vista privilegiado, otra virtud de los modelos bayesianos es que la selección natural puede haber dotado a un organismo con una máquina de análisis estadístico de alta potencia sin tener que instalar un homúnculo matemático en un cargo especial dentro de la estructura burocrática. Se trata, si lo preferís, de estructuras de generación de expectativas con una remarcable competencia que no necesitan comprender. En obras anteriores he prevenido contra postular algún «tejido milagroso» en el cerebro para salvar la papeleta que tiene que salvarse de algún modo, pero sin explicar cómo este tejido milagroso consigue ser tan extraordinario. Estas estructuras bayesianas no tienen nada de mágicas; funcionan perfectamente cuando se ejecutan en computadoras comerciales bien vulgares. Los detalles de cómo se ponen en práctica, a escala neuronal, estas redes jerárquicas de codificación predictiva aún son poco claros, pero no parecen estar fuera de nuestro alcance.[5]

Todo esto está muy bien para la cognición en animales; varios analistas del elegante estudio de Clark aventuran la conclusión tentativa de que las mentes animales son máquinas bayesianas de expectativas que se basan en las probabilidades extraídas de las impresiones pasajeras para guiar su comportamiento. Pero si es así, entonces volvemos a encontrarnos con que las mentes animales son colonias de termitas, no diseñadores inteligentes. Es cierto que son tan competentes que podemos admitir todo tipo de comprensión conductual, pero así y todo, de algún modo importante que aún está por formular, carecen de uno de nuestros Trucos: poseer razones y no solamente actuar por las razones. Los anticipadores bayesianos no necesitan expresar ni representar las razones que rastrean; al igual que la evolución, separan a ciegas el grano de la paja informacional y actúan en consecuencia. Las razones no son cosas que pertenezcan a sus ontologías, no son elementos relevantes de sus imágenes manifiestas.

Pero para nosotros, las razones sí que son cosas. Son los utensilios y los objetos del diseño inteligente de arriba abajo. ¿De dónde proceden? ¿Cómo se instalan en nuestros cerebros? Por fin, creo que puedo argumentar con cierto detalle que nos llegan gracias a la evolución cultural, un proceso nuevo de I+D de menos de un millón de años de antigüedad que diseña, propaga e instala miles y miles de recursos de pensamiento en nuestros cerebros (y solo en los nuestros) y los convierte en mentes, mentes «de verdad».

¿NEURONAS SALVAJES?

Pronto dejaremos atrás la escala de las neuronas, pero antes de hacerlo no puedo resistirme a una especulación que enlaza algunos de nuestros temas y que puede que tenga un papel en la explicación del hecho de que Homo sapiens es, hasta el momento, la única especie de la Tierra con mentes «de verdad», mentes enculturadas repletas de recursos de pensamiento. He sostenido que las neuronas son las descendientes domesticadas de eucariotas muy antiguas que durante eones se enfrentaron a la vida como microorganismos autónomos. Para distinguir a estas eucariotas de las bacterias y de las arqueas, a menudo se las

denominaba protistas, del reino Protista. Sin duda, se las apañaron bien, o no estaríamos aquí, y las protistas actuales, como las amebas o las algas unicelulares, siguen prosperando, gracias a sus muchos y variados talentos. Cuando las protistas empezaron a formar colonias y, finalmente, organismos pluricelulares, se llevaron consigo sus genomas, las instrucciones de todas las competencias que habían ido adquiriendo durante miles de millones de generaciones gracias a la selección natural. En su nuevo ambiente protegido, muchos de estos talentos ya no eran necesarios, de manera que se perdieron; o, mejor dicho, se perdieron las expresiones de los genes relevantes, puesto que no compensaba. Pero resulta más barato y rápido dejar los genes en su sitio y desactivar su expresión durante el desarrollo (o al menos parte de ella) que eliminar totalmente esos genes (si bien es cierto que, si nunca se usan, al final sí que se eliminarán).

He aquí un bonito paralelismo con una táctica empleada en el desarrollo de software: el código heredado. Cuando se revisa un programa (cosa que sucede a menudo; ahora mismo estoy escribiendo este libro en Word 14.4.8, cuyo nombre ya revela su genealogía: la octava revisión menor de la cuarta revisión principal de la decimocuarta versión nueva y mejorada), los programadores serían estúpidos de descartar y eliminar todo el duro trabajo de los programadores anteriores (¡quién sabe cuándo puedes necesitarlo para algo!) y por eso la práctica habitual es dejar intacto el código obsoleto junto al código que lo sustituye y marcarlo como comentario, entre corchetes, asteriscos, o alguna otra convención que el compilador reconozca como tal, de modo que no esté representado en el código que ejecutará la computadora.[6] Recordad que en los genomas no hay distinción entre código fuente y código compilado, pero este mismo silenciamiento del código se puede producir mediante una simple mutación en un gen regulador (un gen que controla la «expresión» de los genes en proteínas).

Cuando se domestican animales, los domesticadores seleccionan, voluntaria o involuntariamente, rasgos apetecibles. Es probable que las diferencias visibles en una camada (o en una manada) sean resultado de diferencias en la expresión de un gen compartido por todos sus individuos. Los domesticadores no necesitan tener ninguna comprensión de los efectos que producen en los genes de su ganado (ni ninguna intención de que se produzcan), pero silenciar ciertos genes sin modificarlos es una ruta prometedora para obtener los resultados

deseados. Así, cuando los descendientes de los animales domesticados escapan y se vuelven silvestres, todo lo que tiene que pasar, a nivel genético, es restaurar las características salvajes que se habían silenciado, sin más que eliminar los «corchetes» que delimitan el código heredado para que vuelva a activarse. Esto puede explicar la rapidísima reversión a los rasgos salvajes que se observa en estirpes asilvestradas. Por ejemplo, en unas pocas generaciones, los cerdos salvajes recuperan muchos rasgos de sus parientes, los jabalíes, tanto en aspecto como en comportamiento, mientras que los caballos asilvestrados, o cimarrones (llamados mustangos en Estados Unidos), están solo a unos cientos de años de sus ancestros domesticados, pero difieren bastante en su temperamento.

La neurona corriente parece lo bastante dócil para pasar toda su vida haciendo el mismo trabajo, sin dejar de tener una chispa de autonomía, una modesta capacidad e inclinación para mejorar su situación si surgen las oportunidades. Podemos imaginar sin dificultad que algunos entornos en los que se encuentran las neuronas favorecen un comportamiento exploratorio más emprendedor y arriesgado, engendrando neuronas más versátiles y agresivas. En estas condiciones, la selección puede favorecer la renovada expresión de genes silenciados durante mucho tiempo, como mínimo en algunas subpoblaciones de neuronas en zonas cruciales del cerebro. Estas neuronas serían, a todos los efectos, neuronas salvajes; un poco menos estables, más egoístas, un poco más propensas a adaptarse a nuevas conexiones con sus vecinas.

¿Qué condiciones podrían favorecer esta evolución? ¿Quizá la llegada de invasores? Las entidades nuevas necesitan apoyos locales para impulsar su propia replicación y ayudarlas a luchar entre ellas para lograr la hegemonía en el cerebro. ¿Qué hace un ejército invasor para intentar ejercer el control del territorio acabado de conquistar? Abre las cárceles y libera a los presos para formar fuerzas de seguridad con nativos curtidos y buenos conocedores del terreno. Esta idea, reutilizada muy a menudo, podría ser la motivación flotante de un desarrollo en paralelo en los cerebros. Estamos a punto de iniciar una descripción detallada y una defensa de los memes, virus de información que controlan hábitos infecciosos. Cuando se expanden por una población, necesitan recursos neuronales para replicarse, del mismo modo que los virus necesitan asumir el control de la maquinaria de copia de las células para crear nuevas copias de sí mismos. Está claro que los virus no pueden

comprender la motivación de esta misión, su naturaleza tiene este efecto en las células con las que se topan; al fin y al cabo, son macromoléculas, no seres vivos. Y los memes, a modo de virus de la mente compuestos de nada más que información, tienen que meterse en las mentes y luego tienen que repetirse, repetirse y repetirse, pero no necesitan comprender nada de todo esto, ni de ninguna otra cosa.

Necesitaremos unos cuantos capítulos para formular y defender esta curiosa idea de elementos informacionales ciegos que incitan a que neuronas rebeldes se alíen en su apoyo, esta idea de que los memes infectan los cerebros.[7] Debo puntualizar que si se produjo algo parecido a una invasión, sin duda debía implicar un proceso coevolutivo en que los cerebros humanos (los únicos cerebros que hasta la fecha han resultado infectados gravemente por los memes) fueron seleccionados para manipular memes, protegiéndolos y ayudándolos en su reproducción, de forma no muy diferente a como las higueras han coevolucionado con las avispas parasitarias que, a cambio, les ayudan en la polinización. Las neuronas presentan muchas formas y tamaños, pero un tipo, las neuronas de Von Economo o neuronas fusiformes, solo se halla en animales con cerebros muy grandes y relaciones sociales complejas: humanos y otros grandes simios, elefantes y cetáceos (ballenas y delfines). Como todas las otras especies que comparten ancestros comunes con estas carecen de neuronas de Von Economo, podemos asumir que estas neuronas fusiformes son un producto de la evolución, reciente e independiente (evolución convergente), aparecido para tener papeles similares. Quizá son solo elementos de comunicación a larga distancia para estos grandes cerebros, pero también se observa que suelen concentrarse en zonas del cerebro implicadas, de modo muy general, en la supervisión propia, la toma de decisiones y la interacción social. La posibilidad de que su presencia en estas estirpes diferentes con comportamientos sociales sofisticados no sea fruto de la coincidencia sugiere que son un candidato probable para su explotación por parte de los memes invasores. Sin duda, puede ser que descubramos otros candidatos, ahora que ya sabemos mejor qué tenemos que buscar.

Ya hemos preparado el terreno para la invasión. Los cerebros son computadoras, pero muy diferentes de las computadoras que usamos actualmente. Formados por miles de millones de neuronas individualizadas que han evolucionado para arreglárselas por sí solas, la arquitectura funcional del cerebro se parece más a un mercado libre que

a una estructura de politburó en que todas las tareas se asignan desde lo alto del escalafón. Es probable que la arquitectura básica de los cerebros animales (incluidos los humanos) esté formada por redes bayesianas que son generadores de expectativas muy competentes y que no tienen que comprender lo que están haciendo. La comprensión (nuestro tipo de comprensión) solo es posible con la reciente llegada de un nuevo tipo de replicador evolutivo, unas entidades informacionales transmitidas culturalmente: los memes.

9

PALABRAS Y EVOLUCIÓN CULTURAL

LA EVOLUCIÓN DE LAS PALABRAS

La supervivencia o preservación de determinadas palabras favorecidas en la lucha por la existencia es selección natural.

C

HARLES DARWIN, El origen del hombre

En el capítulo 7 hemos anticipado el tema de este capítulo con las figuras 7.4 y 7.5, los espacios darwinianos en que podemos representar algunas dimensiones de la evolución cultural. La figura 7.4 (p. 138) muestra las palabras, al igual que los virus, como cosas comparativamente simples (respecto a los organismos o las religiones, por ejemplo) que no se limitan a fluctuar (como la concentración de oxígeno en la atmósfera, como la confianza), sino que se replican: tienen descendencia. La figura 7.5 (p. 139) muestra la evolución de la evolución de la cultura, desde los procesos profundamente darwinianos (con una comprensión mínima y una creación de abajo arriba de productos nuevos mediante procesos de búsqueda aleatorios) hasta los procesos de diseño inteligente (con comprensión y una creación de arriba abajo de productos nuevos mediante procesos de búsqueda dirigidos). Ha llegado el momento de ofrecer y argumentar los detalles de esta evolución. Sostendré que las palabras son el mejor ejemplo de memes, elementos transmitidos culturalmente que evolucionan por replicación diferencial (es decir, por selección natural).

Hay otras especies que muestran ciertos rudimentos de evolución cultural. Los chimpancés tienen algunas tradiciones, como partir frutos secos con piedras, pescar hormigas con ramas o pajitas, expresiones de cortejo y unas pocas más, que no se transmiten genéticamente entre generaciones, sino que son formas de comportarse que dependen de la percepción que tiene la descendencia del comportamiento de los mayores. Las aves tienen diversas formas, bien estudiadas, de adquirir los cantos propios de su especie; las vocalizaciones de las gaviotas y de los gallos y gallinas, por ejemplo, son totalmente instintivas y no requieren de ningún modelo acústico para manifestarse, mientras que los polluelos de la mayoría de otras especies lo que tienen es un «instinto de aprender», en palabras del etólogo Peter Marler, y necesitan oír el canto de sus padres. Varios estudios con intercambio de crías, en que se colocan huevos de una especie en los nidos de otra, han mostrado que los polluelos copian el canto de sus padres adoptivos lo mejor que pueden. Se ha demostrado que un amplio abanico de inclinaciones animales (formas de comportarse) que se creía que eran «instintos» transmitidos genéticamente en realidad son «tradiciones» transmitidas de padres a hijos mediante canales perceptuales, no mediante genes (Avital y Jablonka 2000), pero en cualquier caso, en ninguna de estas especies hay más de un puñado de estas formas adquiridas de comportamiento.

Nosotros, Homo sapiens, somos la única especie (hasta la fecha) con una rica cultura acumulativa, y el ingrediente fundamental de la cultura que lo hace posible es el lenguaje. Gracias a nuestra cultura humana acumulativa hemos engullido nuestro planeta, hemos transformado el entorno y hemos puesto en marcha un suceso de extinción que pronto podrá rivalizar con las extinciones masivas de eones atrás si no tomamos medidas para revertir algunas tendencias actuales. Nuestra especie ha mantenido un crecimiento de población sin igual en cualquier otra especie de vertebrados (si dejamos de lado a nuestras gallinas, vacas y cerdos). En los últimos dos siglos, nuestra población ha aumentado de mil millones a más de siete mil millones, y aunque el ritmo de crecimiento está disminuyendo bastante, llegaremos a los ocho mil millones en menos de una década, según las últimas estimaciones de las Naciones Unidas.

En los últimos 50.000 años, nuestros genes no han cambiado mucho, y con toda probabilidad los cambios que han sufrido han sido debidos, directa o indirectamente, a las nuevas presiones selectivas creadas por las innovaciones culturales humanas, como la cocción de los alimentos, la agricultura, el transporte, la religión y la ciencia. La adopción

generalizada de un nuevo comportamiento crea un trinquete de un solo sentido: en el momento en que casi todo el mundo come alimentos cocinados, el sistema digestivo humano evoluciona genéticamente para que ya no resulte práctico (y posteriormente para que ya no resulte posible) subsistir a base de alimentos crudos. En cuanto el transporte empieza a desplazar personas por todo el mundo, primero poblando todas las islas y continentes y luego explorándolos o invadiéndolos, los seres humanos ya no pueden subsistir sin un ejército de defensas contra las enfermedades que llegan con los viajeros.

Uno de los hechos de la vida, tanto genético como cultural, es que las opciones se convierten en obligaciones. Un astuto truco nuevo que brinda a sus usuarios una clara ventaja sobre sus semejantes pronto se propaga hasta quedar fijado, momento en el cual todos aquellos que no lo hayan adquirido están condenados. Si esconderse de los depredadores en un agujero es un comportamiento que salva vidas lo bastante a menudo, con el tiempo tener un agujero donde esconderse deja de ser una excentricidad de unos pocos miembros de la población para pasar a ser una necesidad de la especie, encarnada en un instinto. La mayoría de especies de mamíferos pueden sintetizar su propia vitamina C, pero no así los primates; su creciente dependencia de una dieta a base de frutas los llevó, después de milenios, a perder su capacidad innata de fabricar vitamina C (¡o lo tomas o lo dejas!). Sin vitamina C puedes contraer escorbuto y otras enfermedades, descubrimiento que se produjo gracias a la navegación humana de larga distancia, que tuvo el efecto de aislar poblaciones «de prueba» cuya salud y dieta eran lo bastante importantes para que las armadas y los cirujanos de abordo se ocuparan de ellas. Este descubrimiento, junto a cientos más, condujo, en unos pocos siglos (un abrir y cerrar de ojos, en términos evolutivos), al exquisito y detalladísimo conocimiento científico de la bioquímica del metabolismo humano. Actualmente estamos obligados a ingerir vitamina C, aunque no a ser frugívoros, ya que podemos tomar las vitaminas según nos convenga. Hoy en día no hay ninguna ley que obligue a nadie a tener una tarjeta de crédito o un teléfono móvil, objetos que no hace tantos años eran poco menos que un lujo; actualmente dependemos tanto de ellos que no es necesario que ninguna ley nos obligue a tenerlos. Puede que el dinero en efectivo no tarde mucho en desaparecer, y entonces seremos usuarios por obligación de la tecnología informática de pagos con tarjeta o sus herederos.

En la actualidad, dependemos tanto de las palabras como de la vitamina C, más o menos. Las palabras son la savia de la evolución cultural (¿o quizá deberíamos decir que el lenguaje es la espina dorsal de la evolución cultural? ¿o que las palabras son el ADN de la evolución cultural? Es difícil resistirse a estas y otras metáforas biológicas, pero siempre que tengamos presente que deben desvestirse de implicaciones ilegítimas, representan un papel valioso). Sin duda, las palabras tienen un papel central e indispensable en nuestra explosiva evolución cultural y es por ello por lo que investigar la evolución de las palabras nos servirá de asequible puerta de acceso a las abrumadoras preguntas sobre la evolución cultural y su papel en la conformación de nuestras mentes. Cuesta saber por dónde empezar; es evidente que en la actualidad disponemos de una profusión cultural y también que hubo un tiempo en que nuestros antecesores no tenían cultura, de modo que tuvo que evolucionar de un modo u otro, pero ¿cómo? La respuesta no es obvia en absoluto. Tal como observan Richerson y Boyd (2004):

Es necesario un poco de teorización científica para convencernos de que la existencia de cultura humana es un profundo misterio evolutivo, al mismo nivel que el del origen de la vida. (p. 126)

Darwin empezó sabiamente su «largo argumento» a medio camino. El origen de las especies no dice nada sobre el tema de cómo empezó todo y deja para otra ocasión los detalles más nebulosos sobre el origen último de la vida. Doce años después, en 1871, en una famosa carta a Joseph Hooker, contempló la posibilidad de que quizá las condiciones en una «pequeña charca cálida» serían las justas para que la vida se pusiera en marcha; más de doce décadas después aún estamos dando vueltas a esa hipótesis de la charca cálida y otras posibilidades. El origen del lenguaje humano es otro misterio sin resolver y desde Darwin ha incitado curiosidad y controversia por razones muy parecidas. Para empezar, hay muy pocos «restos fósiles» de sus inicios, y los que se pueden identificar con seguridad están sometidos a interpretaciones diversas. Además, ambos misterios se refieren a sucesos trascendentales que quizá solo hayan ocurrido una única vez en este planeta.

En los dos casos (y aquí hay que ser muy cuidadoso al expresarlo), es posible que la vida y el lenguaje hayan surgido varias veces, o incluso muchas, pero si así fuera, estos otros inicios se han extinguido sin dejar ningún resto que podamos detectar (aunque ¿quién sabe qué nos puede deparar el futuro?). Tiene que haber habido un origen para cada uno, por lo menos, y no hay pruebas convincentes de más de un origen, de modo que empecemos con la suposición de que hubo un único inicio en ambos casos. Ahora nos queda la tarea de elaborar hipótesis científicamente plausibles e indirectamente comprobables sobre cómo han podido darse estas dos historias de exitosa I+D evolutiva. El trabajo empírico en ambos campos ha efectuado avances considerables en las últimas décadas, avances que estimulan investigaciones adicionales que incorporan la suposición predeterminada (y tentativa) de que los «árboles» de los linajes existentes convergen en un único tronco.

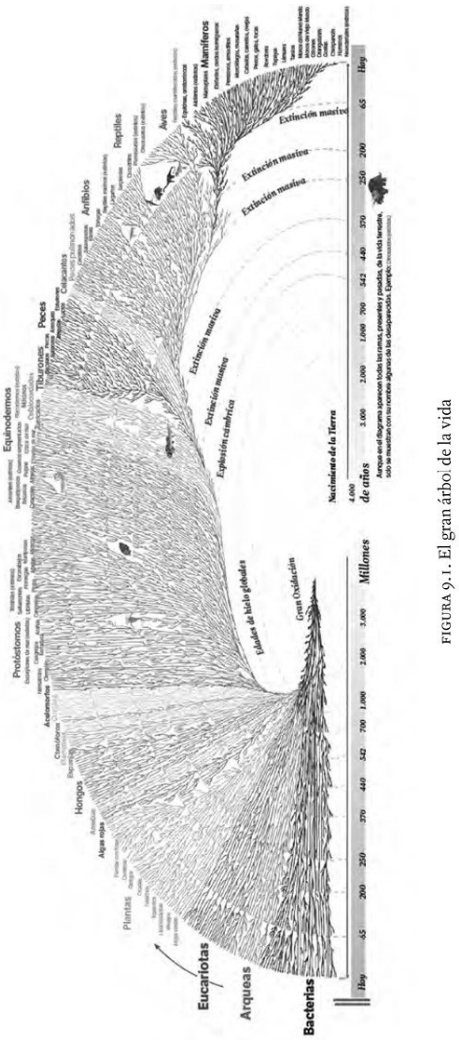


FIGURA 9.1. El gran árbol de la vida

Los diagramas filogenéticos, o cladogramas (como el gran árbol de la vida que se muestra en la figura 9.1), que muestran todas las especies, o bien los árboles más reducidos de la progenie de linajes concretos, se vuelven cada vez más claros gracias a la investigación bioinformática sobre la acumulación de diferencias en secuencias de ADN y, así, se pueden rellenar los vacíos y corregir los errores de los anteriores trabajos detectivescos anatómicos y fisiológicos.[1] Los árboles glosogenéticos, que representan linajes de lenguas (figura 9.2), también son un popular recurso de pensamiento que expone las relaciones de descendencia entre familias lingüísticas (y palabras) a lo largo de los siglos.

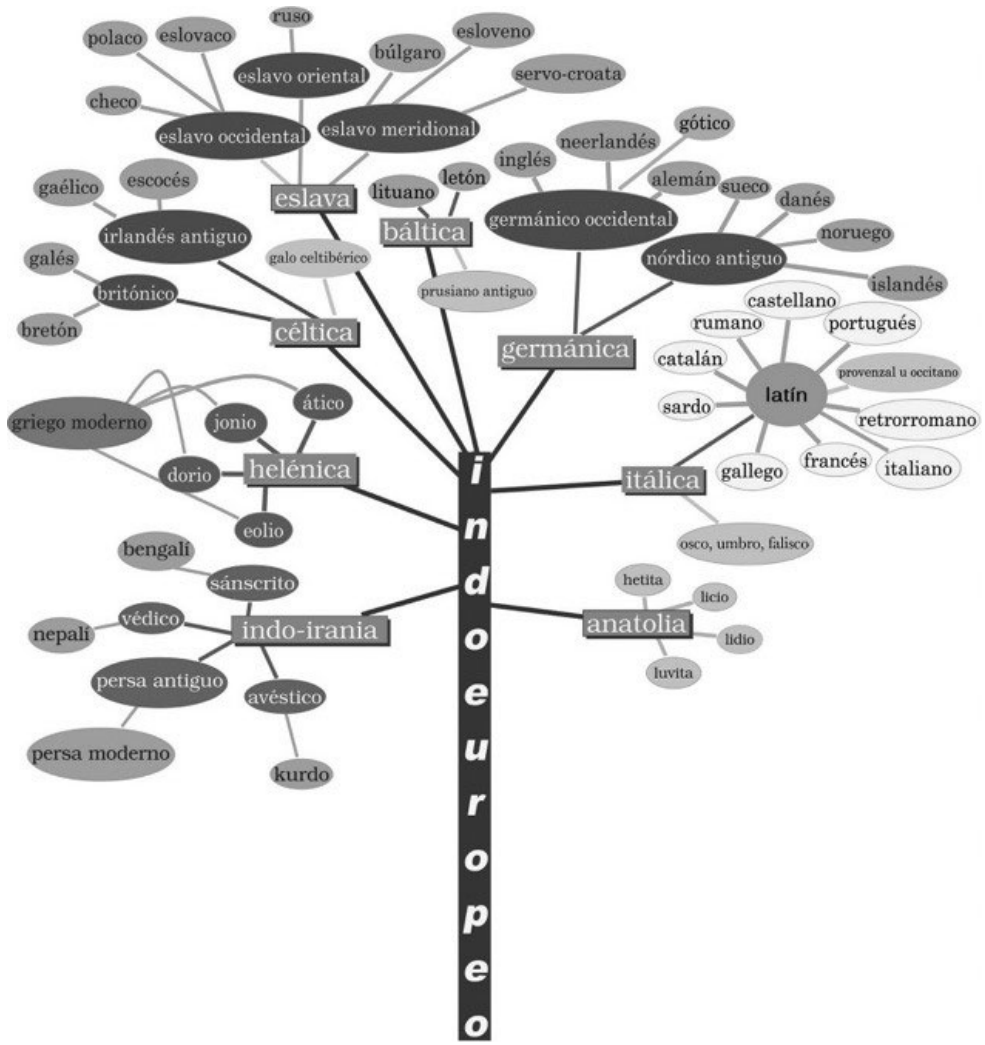


FIGURA 9.2. Árbol glosogenético de las lenguas indoeuropeas

Los biólogos que componen árboles filogenéticos se topan con dificultades cuando tienen que representar las anastomosis, la unión de dos linajes hasta ese momento separados, un fenómeno que ahora sabemos fue muy habitual en las primeras épocas de la vida (basta con recordar el origen endosimbiótico de los eucariotas). Los lingüistas históricos que elaboran árboles glosogenéticos se enfrentan a numerosas anastomosis de lenguas que se unen (creando pidgins y criollos, por ejemplo) y a los saltos (aún más numerosos) de palabras de una lengua a otra. Es evidente que términos inuit como iglú y cayac serán adoptados por muchas otras lenguas en cuanto se familiaricen con los iglús y los cayacs, del mismo modo que fiesta o salsa son palabras castellanas casi universalmente conocidas por personas que no hablan castellano.

Dawkins (2004) señala que, en muchos casos, los árboles que muestran los linajes de genes individuales son más fiables e informativos que los árboles tradicionales que muestran la descendencia de especies, a causa de la transferencia genética horizontal, el fenómeno por el cual hay genes que «saltan» de una especie o familia a otra. Cada vez está más claro que es muy frecuente el intercambio o la compartición de genes mediante procesos independientes de la reproducción (la transferencia vertical de genes), sobre todo en bacterias y otros organismos unicelulares. De modo parecido, las etimologías de las palabras son más seguras que la descendencia de las lenguas en las que se hallan, a causa de la transferencia horizontal de palabras entre idiomas.

La idea de que las lenguas evolucionan, de que las palabras de hoy son, de algún modo, descendientes de las palabras del pasado es más antigua que la teoría de Darwin de la evolución de las especies. Por ejemplo, se sabía que los textos de la Ilíada y la Odisea de Homero descendían, a base de copias, de textos que descendían de textos que descendían de textos que se remontaban a sus antecesores orales de la época homérica. Los filólogos y los paleógrafos han reconstruido linajes de lenguas y de manuscritos (p. ej., las diversas copias conservadas de los Diálogos de Platón) desde el Renacimiento, y algunas de las más recientes técnicas bioinformáticas empleadas hoy en día para determinar relaciones entre genomas son, a su vez, descendientes perfeccionados de las técnicas elaboradas para rastrear patrones de errores (mutaciones) en textos antiguos. Tal como Darwin señaló, «la formación de diferentes lenguas y de distintas especies y las pruebas de que ambos se han desarrollado mediante un proceso gradual son, curiosamente, iguales» (1871). Seguiremos el ejemplo de Darwin y empezaremos a medio camino,

aplazando el origen último del lenguaje hasta el final del capítulo; así, veremos lo que podemos decir con seguridad sobre la evolución actual del lenguaje y, en concreto, de las palabras.

UNA MIRADA MÁS ATENTA A LAS PALABRAS

¿Qué son las palabras? En este momento resulta útil un poco de jerga filosófica y presentar la distinción entre tipo y caso, formulada por primera vez por Charles Sanders Peirce en 1906:

(1) Palabra es una palabra, y en esta frase hay tres casos de esa palabra.

¿Cuántas palabras hay en la frase (1)? Catorce si contamos casos (como el recuento de palabras en Microsoft Word), pero solo doce si contamos tipos. Si pronunciamos la frase (1) en voz alta habrá tres sucesos acústicos diferentes, tres casos del tipo palabra, que se producirán en sucesión durante unos pocos segundos. En la página impresa hay tres diferentes patrones de tinta que se parecen mucho (el primero tiene mayúscula inicial y está en cursiva, pero sigue siendo un caso del tipo palabra). Los casos pueden ser sucesos audibles, patrones de tinta, patrones de publicidad aérea, muescas en una piedra o cadenas de bits en computadoras.[2]

También pueden ser sucesos silenciosos en el cerebro. Cuando leemos la frase (1) en silencio, además de los casos visibles en la página, hay casos en nuestro cerebro que, como los casos expresados en voz alta, tienen propiedades físicas y momentos cronometrables de aparición (los casos escritos pueden tener una vida muy larga, desde el momento en que la tinta todavía no está seca hasta que el papel se pudre o se quema). Aún no sabemos cómo identificar casos cerebrales de las palabras por sus propiedades físicas (todavía no sabemos cómo hacer «lecturas del cerebro», aunque estamos cada vez más cerca) pero podemos estar seguros de que los casos cerebrales del tipo palabra serán físicamente tan

diferentes de los casos escritos y hablados como estos lo son entre sí. Los casos cerebrales no tendrán el aspecto de palabra ni sonarán como palabra (son sucesos cerebrales y en el cerebro todo es muy sombrío), pero sin duda serán similares físicamente a algunos de los sucesos que tienen lugar en el cerebro cuando ves u oyes palabra.

Los teóricos de la conciencia suelen pasar por alto este curioso hecho (como veremos en el capítulo 14), con unos efectos devastadores. Leed la frase (1) en silencio pero con voz de falsete. Ahora leedla en silencio pero con acento francés (con muchas erres guturales y acentuando a final de palabra). ¿Os ha «sonado francés» en el oído de vuestra mente? Pronunciar una palabra en la mente es trabajo añadido, algo innecesario para un usuario habitual de la lengua, pero un apoyo valioso al aprender un nuevo idioma (como mover los labios al leer). El cerebro está equipado con una maquinaria auditiva que se especializa en identificar sonidos del habla y con una maquinaria articuladora que se especializa en controlar los labios, la lengua y la laringe. Estos sistemas neuronales se forman durante el aprendizaje de un idioma y tienen un papel relevante en el proceso de «decir las palabras en tu mente», pero esta activación de zonas del cerebro no produce ningún sonido (no busquéis ninguna vibración que pueda ser captada por un micrófono sensible implantado en el cerebro). Ahora cerrad los ojos y «visualizad» en el ojo de vuestra mente las primeras cuatro palabras de la frase (1), primero en amarillo sobre negro y luego en negro sobre blanco. Si juzgamos a partir de la introspección (un método delicado y poco fiable, pero indispensable en momentos como este, al iniciar la investigación), cuando «nos pasan palabras por la cabeza» no solemos complicarnos la vida «oyéndolas» con toda la parafernalia de rasgos auditivos como el tono, la correcta pronunciación o la entonación, ni «viéndolas» en algún tipo, tamaño y color de letra concretos. Cuando no puedes quitarte su nombre de la cabeza, ¿está siempre en mayúsculas? ¿Pronunciado con un suspiro?

Así pues, parece que hay gran cantidad de ejemplos intermedios de palabras (palabras específicas, concretas) «casificadas» (es decir, convertidas en casos) en nuestras mentes, sin necesidad de distinguir si son habladas o escritas, oídas o vistas. Y también parece haber un pensamiento «sin palabras» en que no nos preocupamos de «encontrar» todas las palabras, sino que solo «casificamos» su significado más descarnado. En el fenómeno «en la punta de la lengua», por ejemplo, podemos saber mucho sobre la palabra que estamos buscando, pero somos incapaces de recuperarla hasta que, de golpe, «nos viene a la

cabeza». Empieza con efe, tiene cuatro sílabas y significa algo como ‘ir a buscar comida para los caballos’… ¡Ah, ya lo tengo!: forrajear. A veces, la imposibilidad de hallar la palabra que estamos buscando a tientas nos bloquea, nos impide analizar detalladamente la solución de un problema; en otros casos, en cambio, podemos ingeniárnoslas sin palabras, con significados desnudos, con ideas que no tienen que estar en castellano ni en francés ni en ningún otro idioma. Aquí surge una cuestión interesante: ¿podríamos hacer esto si nuestros sistemas neuronales, nuestra «disciplina mental» no se hubiera entrenado mediante el aprendizaje de nuestra lengua materna? ¿Podría ser el pensamiento sin palabras una especie de esquí acuático sin esquís, que depende de un engranaje que puedes poner en marcha una vez ya has arrancado? Los psicolingüistas y los neurolingüistas han empezado con buen pie al complementar (y, a menudo, sustituir) esta introspección informal con experimentos controlados y modelos de generación y percepción del habla (Jackendoff 2002, especialmente el cap. 7 «Implicaciones para el procesamiento», es una introducción estimable).

El problema con la introspección es que acepta la ilusión de que hay un ojo interior que ve y un oído interior que oye (y una mente interior que piensa) estos elementos íntimamente familiares, los objetos de la conciencia que desfilan por el escenario del Teatro Cartesiano (Dennett 1991). No hay ningún Teatro Cartesiano, solo parece haber un Teatro Cartesiano. En efecto, parece que hay estos elementos, audibles, visibles y pensables y (dejando de lado la magia) no podrían parecer así si no hubiera realmente casos físicos en el cerebro interpretando estos papeles; ahora bien, cómo algo en el cerebro consigue interpretar cualquiera de estos papeles es un tema de investigación científica futura, no un tema para la introspección.

Estamos yendo demasiado rápido; en el capítulo 14 ya tendremos ocasión de reunir las implicaciones de estas observaciones para una teoría de la conciencia. De momento, solo quiero poner de manifiesto el hecho evidente de que además de casos públicos de las palabras, hay casos privados e internos de estas y hacer notar el hecho, mucho menos evidente, de que aún no sabemos gran cosa acerca de sus propiedades físicas. Los casos internos parecen asemejarse a casos externos, pero esto es porque emplean los mismos circuitos neuronales que usamos para detectar las semejanzas y diferencias entre los casos externos, no porque estos circuitos neuronales reproduzcan copias de lo que identifican.[3]

Aunque algún día lleguemos a dominar el arte de identificar las palabras que las personas contemplan (casifican) en sus cerebros, esta capacidad no sería, por sí misma, una «lectura de la mente» en el sentido de descubrir sus creencias ni tan solo sus pensamientos. Para demostrarlo, basta con pronunciar cinco veces en silencio «¡Muera la democracia!»; ahora supongamos que puedo leer tus casos cerebrales, de modo que podría descubrir lo que estás diciendo a escondidas; ¿me mostraría eso que crees en lo que estás diciendo? No, ¿verdad? Quizá una lectura de mentes de este tipo pueda ser posible en el futuro; en tal caso, la identificación de casos cerebrales de palabras equivaldría, más o menos, a las escuchas telefónicas actuales: una indicación, pero ni mucho menos una prueba concluyente para atribuir un determinado pensamiento.

En cualquier caso, hay pocas posibilidades de que vuestros casos cerebrales de palabra sean físicamente parecidos a mis casos cerebrales de palabra (en forma, ubicación u otras propiedades físicas)[4] y tampoco apostaría mucho contra la propuesta de que tenéis muchas clases físicamente diferentes de casos cerebrales de palabra, que implican a agrupaciones de neuronas repartidas por las zonas de procesamiento del lenguaje de vuestros cerebros, más diferentes en forma y otras propiedades físicas que estos casos escritos:

palabra palabra PALABRA PALABRA palabra PALABRA palabra

Lo que hace que algo sea un caso de un tipo no es un simple parecido; el caso escrito «gato» y el sonido que emites al leerlo en voz alta son dos casos del tipo gato que no se parecen en nada. Incluso los casos hablados pueden tener propiedades muy diferentes: un bajo profundo que pronuncie gato crea un sonido que tiene pocas semejanzas físicas con la misma palabra cuchicheada por una niña de cinco años, pero ambos son fácilmente reconocibles por los hablantes de castellano como casos de gato. Supongamos que Alberto dice a Berta chocolate, esta escribe chocolate en un papelito y se lo pasa a Carlos, quien susurra chocolate a Diana, quien escribe chocolate en su iPhone y se lo muestra a Emilia, y así sucesivamente. Cualquier proceso que genere un nuevo caso de un tipo a partir de un caso existente de un tipo cuenta como replicación, tanto si los casos son físicamente idénticos o parecidos como si no. Los casos

de las palabras son objetos físicos de una clase u otra, pero podríamos decir que las palabras están hechas de información, como el software, y que su individualidad, en la mayoría de situaciones, procede de los tipos, no de los casos. La versión de Microsoft Word instalada en vuestra computadora es un caso, pero cuando hablamos de Microsoft Word (incluso Microsoft Word 14.4.1) en general nos estamos refiriendo al tipo, no al caso.

A Darwin le parecía evidente la idea de que las palabras han evolucionado por selección natural y podríamos pensar que Noam Chomsky, quien lanzó la idea de un innato Dispositivo de Adquisición del Lenguaje en el cerebro, miraría con simpatía las explicaciones evolutivas acerca de cómo las lenguas en general y las palabras en particular han llegado a tener los rasgos remarcables que tienen, pero lo cierto es que Chomsky ha denigrado el pensamiento evolutivo en lingüística en casi todos sus aspectos (volveremos sobre esto en el capítulo 12). Tras la estela de Chomsky, muchos lingüistas y filósofos del lenguaje se resisten firmemente al pensamiento evolutivo, con unas pocas excepciones destacables. En su obra pionera Language, Thought and Other Biological Categories (1984), la filósofa Ruth Millikan conjeturó que las palabras eran descendientes de «gesticulaciones del aparato vocal». El filósofo David Kaplan (1990) presentó un modelo de las palabras como prolongaciones y de las expresiones o inscripciones (u ocurrencias en la cabeza) como etapas, un modelo «naturalista» claramente inspirado por Darwin, sobre todo en su rechazo de las esencias platónicas inmutables y en su reconocimiento de que los casos pueden tener propiedades físicas muy diferentes aun siendo del mismo tipo (de hecho, Kaplan despreciaba la terminología de caso y tipo, durante mucho tiempo un pilar de la filosofía del lenguaje, a causa del olor a esencialismo que desprende, pero finalmente transigió: «Aunque yo prefiera expresión e inscripción, no me importa si queréis continuar llamando casos a las expresiones y las inscripciones, siempre y cuando no quedemos aprisionados en la metafísica del modelo de casos y tipos». El coqueteo de Kaplan con la evolución encontró una fuerte resistencia en el grupo de Chomsky en el MIT; el filósofo Silvain Bromberger (2011) reaccionó con el estilo típico del MIT:

El hecho indiscutible es que toda esta cháchara de palabras que cambian es, a lo sumo, un cómodo atajo y una forma de no comprometerse con los

detalles empíricos. Pero la moraleja, una vez más, es que cualquiera que esté seriamente interesado en la ontología del lenguaje no debería tomar las façons de parler sobre el cambio al pie de la letra […] ¡Son las personas las que cambian, no las «palabras»! (pp. 496-497)

Bromberger añadió una nota al pie: «si son entidades abstractas, ¿cómo pueden cambiar?» Como si esto demostrara algo. ¿Qué debe Bromberger pensar sobre los genes? ¿Que también son nada más que una façon de parler? La oposición al pensamiento evolutivo por parte de Chomsky y algunos de sus colegas y seguidores de todo el mundo solía desalentar a los filósofos del lenguaje a considerar las posibilidades, pero los tiempos están cambiando. Además de Millikan, Kaplan y yo mismo, Daniel Cloud (2015) ha asumido este punto de vista y Mark Richard (futura publicación) explora la analogía entre especie biológica y significado lingüístico (sobre lo que volveremos en el capítulo 11).

Mi colega, el lingüista Ray Jackendoff (2002), ha elaborado una teoría del lenguaje fundada en la evolución, que intenta aliarse con la neurociencia reciente y, según su descripción, las palabras son estructuras de la memoria, autónomas en el sentido de que se deben adquirir de forma independiente (deben aprenderse).[5] Son elementos de información tal como los definimos en el capítulo 6. Pero hay otras estructuras informacionales: cuentos, poemas, canciones, lemas, consignas, mitos, técnicas, «prácticas recomendadas», escuelas de pensamiento, credos, supersticiones, sistemas operativos, navegadores web y applets de Java, entre muchos otros. Las estructuras informacionales se presentan en diversos tamaños (atendiendo al número de partes, no al tamaño físico de algunos de sus casos, como en el famoso cartel de HOLLYWOOD en Los Ángeles). Una novela es bastante larga, los poemas suelen ser bastante más pequeños y las señales de tráfico aún más (STOP), mientras que las marcas comerciales a menudo son solamente una única forma, de cualquier material.

Además de las partes audibles o visibles de sus casos, las palabras contienen muchas partes informacionales (lo que las hace sustantivos y verbos, comparativos y plurales, etc.). Las palabras son autónomas, en ciertos aspectos; pueden migrar de un idioma a otro y darse en muchos papeles diferentes, públicos y privados. Una palabra, como un virus, es una clase mínima de agente: quiere ser dicha (Dennet 1991). ¿Por qué?

Porque si no es así, pronto desaparecerá. Una palabra es egoísta en el mismo sentido en que un gen es egoísta (Dawkins 1976). Este uso metafórico ha resultado ser muy efectivo para guiar nuestras mentes hacia perspectivas esclarecedoras de la evolución (y no tengáis miedo de alguna metáfora de vez en cuando, no os morderá, pero debéis tener claro que sabéis cómo canjearla por los hechos lisos y llanos cuando sea necesario).

Un objeto informacional no tiene ninguna mente, por supuesto; no más que un virus, pero al igual que un virus, está diseñado (por la evolución, en su mayor parte) para provocar y mejorar su propia replicación. Cada caso que genera es un individuo de su descendencia. El conjunto de casos que descienden de un caso ancestral forma un tipo, que, de este modo, se asemeja a una especie biológica. Ahora podemos entender la reticencia de Kaplan a mantener la distinción entre casos y tipos, ya que los casos se desdibujan gradualmente hasta que surge un nuevo tipo, de modo no muy diferente a cómo un linaje de descendientes de algún dinosaurio llegó a ser, con el paso del tiempo, un conjunto de miembros de una nueva especie de ave. Algunos descendientes de un caso de una palabra serán expresiones privadas; su anfitrión humano habla consigo mismo, quizá repitiendo de forma obsesiva la palabra en su mente, provocando una explosión poblacional de casos que, así, se construyen un nicho aún más sólido en un cerebro (y puede ser que muchas más casificaciones, o descendientes, internas aparezcan sin nuestra atención consciente; en este mismo momento, puede que las palabras estén replicándose competitivamente en tu cabeza, tan inadvertidas como los microbios que se reproducen en tu estómago; volveremos a ello más adelante). Algunos descendientes se expresarán en voz alta, se escribirán o incluso se imprimirán en libros; unos pocos se alojarán en otros cerebros, donde hallarán un hogar ya preparado para ellos (si sus anfitriones los reconocen) o tendrán que hacerse un nuevo hueco.

¿CÓMO SE REPRODUCEN LAS PALABRAS?

Como pasa con los virus, cuanto más mejor. Una única copia de un virus puede infectar a un nuevo anfitrión, pero una riada de virus tiene muchas más posibilidades de establecer una buena cabeza de puente y, así, una especie de colonia reproductora. Del mismo modo, oír una palabra una

sola vez puede ser suficiente para establecer un nuevo elemento de vocabulario, un nuevo generador de casos de palabras, sobre todo para un oyente adulto interesado, pero lograr una impresión en un niño es más probable si la palabra se oye varias veces.

¿Cómo se instalan las palabras en los cerebros de los niños? En promedio, los niños aprenden unas siete palabras al día, desde su nacimiento hasta los seis años (¿cómo lo sabemos? Fácil: medimos el vocabulario de un niño de seis años, que son unas 15.000 palabras, y lo dividimos por el número de días que ha estado vivo: 2.190. Aprende unas 200 palabras durante los primeros dos años de vida y luego el proceso se acelera durante unos cuantos años, antes de disminuir bruscamente… ¿cuántas palabras habéis aprendido esta semana?). Si partimos de sus primeros días de aprendizaje de palabras, descubriremos que, en promedio, se necesitan seis casificaciones de una palabra en presencia del niño para que produzca sus primeros esfuerzos claros de decir esa palabra, para expresar una copia (Roy 2013; véase también la brillante charla TED de Roy en http://www.ted.com/talks/deb_roy_the_birth_of_a_word). Se podría decir que, a diferencia de un virus, una palabra ha de tener varios padres antes de que pueda nacer, pero no tienen que estar todos juntos al mismo tiempo. La mayoría de palabras que oyen los bebés no se dirigen a ellos; los bebés están continuamente escuchando el habla de sus padres y cuidadores. La primera ocurrencia de una palabra es simplemente un nuevo suceso auditivo en algún contexto perceptual complejo y, en su mayor parte, indescifrable, pero de algún modo provoca una impresión en el cerebro. La segunda ocurrencia se añade a esa primera impresión (si el bebé pudiera hablar, diría «Oh, es ese sonido otra vez») y se da en un contexto que puede compartir algunas características con el primero (o no). La tercera ocurrencia es un pelín más familiar y el contexto en que se da puede empezar a tomar forma. La cuarta, quinta y sexta ocurrencias permiten captar la firma auditiva, lo que los lingüistas llaman fonología, y crean una especie de anclaje en el cerebro. En este momento, el niño tiene un objetivo para pasar a la acción: ¡decirlo! Podemos suponer que la disposición para formar este objetivo, a estas alturas, se ha instalado genéticamente en el genoma humano, pero en los primeros días de la vocalización y comunicación de los homininos[6] es posible que fuera una particularidad variable, una ampliación fortuita de una tendencia general a favor de la imitación y la interacción con padres y cuidadores (en el capítulo 12 hablaremos mucho más acerca del origen del lenguaje).

En cuanto el niño empieza a decir palabras, los adultos responden hablando más lento y simplificando la vocalización cuando están en su compañía (Roy 2015). El niño ya no es tratado como un oyente involuntario, sino como un candidato a participar en la conversación. Tras unas pocas generaciones de casos, y con las reacciones suscitadas ante la ejecución de sus propios intentos, el niño empieza a producir copias reconocibles, las primeras palabras del niño. Pero el niño ni tiene por qué entenderlas, ni tan solo entender que son palabras; ha desarrollado un hábito de expresar estos sonidos estructurados en contextos cada vez más específicos, y a veces ello aporta una recompensa, igual que el lloro instintivo tiende a aportar comida, caricias o comodidad. Los sonidos de las palabras concretas se vuelven cada vez más familiares, reconocibles e identificables, a medida que se acumulan las repeticiones, tanto de expresión como de audición. En ese momento, una palabra pronunciable se ha instalado en el cerebro del bebé, y ya sirve para… ¡nada! Para nada excepto para reproducirse a sí misma. Pronto el bebé comienza a adquirir usos discernibles (inconscientemente) de la palabra y, poco a poco, empieza a significarle algo.

Estas generaciones de descendientes pueden ser inconscientes, ya que es muy posible que el bebé aún no haya desarrollado una conciencia competente a partir de su riada de estímulos y reacciones. Estamos ante un tema polémico que hay que abordar con cautela. Ahora no es necesario que tracemos una frontera clara entre actividades conscientes e inconscientes, incluso aunque, después de todo, sí que exista una línea bien dibujada. Igual que cuando nos despertamos de un sueño profundo lo hacemos gradualmente, sin que haya ningún momento que podamos identificar como el primer momento consciente, el desarrollo de la conciencia de un bebé, en cualquier forma digna de tal nombre (algo más que una simple sensibilidad y respuesta a los estímulos, que se halla también en plantas y en bacterias), es probablemente gradual, como todas las grandes transformaciones de la vida.

Algunas personas se aferran a la idea de que la conciencia es la gran excepción, una propiedad a todo o nada que divide el universo en dos clases totalmente disjuntas: las cosas para las que hay algo que es ser como esa cosa, sentirse como esa cosa, y aquellas para las que no lo hay (Nagel 1974; Searle 1992; Chalmers 1996; McGinn 1999). La verdad es que no he encontrado nunca ningún argumento que me convenza de que tendría que ser así. Se me antoja un sospechoso descendiente del vitalismo, la doctrina hoy en día casi abandonada de que estar vivo es una

propiedad metafísica especial que requiere una cierta infusión de un algo maravilloso, un élan vital. Como este punto de vista sobre la conciencia, metafísicamente extravagante, aún es popular entre personas sensatas, entre las cuales se cuentan filósofos y científicos, en este punto admitiré que los bebés solo son probablemente no muy conscientes, no realmente conscientes de la descendencia de palabras que empieza a medrar en sus cerebros. Si estas personas tan razonables algún día demuestran que están en lo cierto (que realmente existe una división cósmica), ya revisaré mi afirmación. Si mañana se consigue identificar esta chispa mágica, aceptaré que la conciencia llega al bebé (o al feto) en el Momento Glorioso M, como sabemos, pero a continuación insistiré en que la conciencia no desarrolla su conjunto habitual de talentos, la capacidad del agente consciente de hacer cosas porque es consciente, hasta que ese agente es ocupado por miles de memes (no solo palabras) que (re)organizan las conexiones neuronales de las que dependen esos talentos.

Estamos observando ese breve periodo en la vida de un bebé en que vocalizaciones aleatorias (o quizá intentos de copiar un sonido oído) pueden dar lugar a palabras infantiles sin sentido que nunca ascenderán más allá de este desarrollo sin función. Estos microhábitos vocales se ensayan durante meses hasta que o bien se extinguen frente a competidores con vocalizaciones más eficaces o bien adquieren un significado local gracias a interlocutores que cooperan con entusiasmo. Para la mayoría de personas, sus palabras infantiles son las únicas que habrán «acuñado» con éxito a lo largo de la vida. Si sus padres y hermanos le repiten estas palabras sin sentido, sin asignarles nunca ningún significado, serán cómplices en la creación de un meme genuino y pronunciable sin ninguna otra función que reproducirse, un virus mental comensal que ni ayuda ni perjudica, pero que prospera durante un tiempo.[7]

Las palabras más eficaces, los memes mutualistas que se convierten en el equipo o vocabulario indispensable de cualquier persona, adquieren su semántica, su sintaxis y sus connotaciones mediante un proceso coevolutivo gradual, casi darwiniano, de aprendizaje de abajo arriba, que se aprovecha de las capacidades de detección de patrones del cerebro y de su conjunto prediseñado de detectores de ofrecimientos (Gorniak 2005; Gorniak y Roy 2006). Esta escueta afirmación puede parecer una especulación irresponsable a muchos lingüistas teóricos, teniendo en cuenta toda la complejidad de la semántica y la sintaxis de las lenguas

naturales; en realidad, la afirmación es muy modesta: en primer lugar llega la fonología, que crea un nodo o un centro en el cerebro para la firma auditiva de la palabra; esto se convierte en la base, el amarre, el punto de recogida para que la semántica y la sintaxis se desarrollen alrededor del sonido, junto con su perfil articulatorio (el cómo decirlo). No voy a tomar partido en las controversias sobre el tema, a pesar de que tengo mis convicciones; ya tengo bastante que defender para tener que respaldar también algún veredicto sobre cuánta tendencia, cuánta «gramática universal», para usar el término chomskiano, tiene que estar instalada genéticamente en el dispositivo de adquisición de lenguaje, ni tampoco voy a especificar de ningún modo qué forma debería tener o podría tener esta tendencia. De una manera u otra, sin ninguna «construcción teórica» deliberada (y mucho menos conciencia), el cerebro del bebé se centra en hábitos que compartirá con sus padres, respondiendo con sensibilidad a la información semántica que se le presenta a sus sentidos.[8]

Actualmente, la instalación de una lengua materna en un bebé humano es un proceso bien diseñado, que sin duda se beneficia de los impactos selectivos de miles de generaciones de aprendices del lenguaje humano. Durante el camino, los procesos evolutivos han descubierto muchos atajos para facilitar la adquisición de las lenguas, la pronunciación y el reconocimiento de palabras y la formulación y la expresión de afirmaciones, preguntas, peticiones y órdenes. Inicialmente, estos procesos evolutivos no habrían implicado una replicación diferencial de genes, sino más bien la mucho más rápida replicación diferencial de memes. El lenguaje evolucionó para adaptarse al cerebro antes de que el cerebro evolucionara para acomodarse mejor al lenguaje.[9]

Parece claro que los memes pronunciables están, hasta cierto punto, constreñidos por la física y la fisiología de la articulación y el oído humanos. Por mucha variación que haya en una lengua particular, los niños adquieren la competencia adulta con el reconocimiento y la articulación y sin apenas instrucción. Sin duda, las propiedades semánticas que comparten sus recién acuñadas palabras, una vez sólidamente instaladas, dependen mucho de las estructuras de la parte no lingüística del Umwelt humano, la imagen manifiesta de ofrecimientos y las acciones que se ocupan de ellos. En esto, los niños también adquieren la semántica sin apenas formación. En muchas culturas, los padres gustan de enseñar los nombres de las cosas a sus hijos, pero por muy entusiastas que sean, los hijos adquieren los significados de la mayoría de las

palabras poco a poco, sin una atención explícita y sin ninguna orientación parental deliberada. La exposición repetida a miles de palabras en diferentes contextos proporciona casi toda la información que necesitan para llegar a un sentido fiable de qué es lo que deben de significar, con solo unos pocos ajustes posteriores cuando la ocasión lo demanda (¿Es un poni un tipo de caballo? ¿Es un zorro un tipo de perro? ¿Estoy avergonzado o arrepentido?).

Sócrates estaba perplejo por el hecho de que la gente pudiera encontrar la definición de una palabra sin más que discutir sobre ella unos con otros. Si no sabían ya qué significaba la palabra, ¿de qué les sirve preguntarse unos a otros? Y si ya sabían lo que significaba, ¿por qué preocuparse en definirla? ¿Y por qué debería ser tan difícil definirla? Parte de la solución a este dilema es darse cuenta de que comprender una palabra no es lo mismo que haber adquirido una definición de ella.[10]

Quizá, además de estas restricciones semánticas y fonológicas haya restricciones sintácticas específicas, genéticamente diseñadas, y atractores impuestos por las propiedades estructurales de nuestros cerebros. Acaso son accidentes fosilizados de nuestra historia evolutiva que ahora limitan la clase de lenguajes que se pueden aprender, de una forma que podría haber sido diferente si la historia hubiera tomado un curso ligeramente distinto. Un ejemplo nos puede ayudar a imaginar esto: supongamos, como hacen algunos teóricos, que el lenguaje haya tenido su origen en la gestualidad, no en la vocalización; podría ser que algún principio de ordenación que resultara óptimo para un lenguaje gestual (por ejemplo, porque minimizaba cambios súbitos en la posición de los brazos, que hubiera forzado en demasía los músculos de gestualizadores locuaces) hubiera sido arrastrado como restricción hasta los lenguajes vocalizados, no porque los hiciera más eficientes, sino porque antes solía hacer más eficientes las gestualizaciones del lenguaje. No es un caso muy diferente de la distribución de letras en los teclados QWERTYUIOP: en su momento era una buena razón, cuando las máquinas de escribir mecánicas tenían palancas para las teclas y para los tipos; las combinaciones de letras muy frecuentes en inglés, como «th» o «st» se asignaron a teclas muy separadas para evitar que los tipos se bloquearan al presionar sus teclas en rápida sucesión. Quizá las restricciones sintácticas impuestas por las estructuras del cerebro son óptimas globalmente; dada la arquitectura neuronal de los cerebros de los mamíferos en la Tierra en el momento en que evolucionó el primer lenguaje, no sería sorprendente que los lenguajes terrestres tuvieran estos

rasgos estructurales. O acaso son Buenos Trucos cuya aparición podríamos predecir en cualquier lenguaje natural en cualquier rincón del Universo; si los trucos son tan buenos siempre y en todas partes, es factible que, a estas alturas, se hayan instalado en nuestro genoma mediante el efecto Baldwin (es decir, que una innovación conductual X demuestra ser tan útil que cualquier individuo que no tiene X está en situación de desventaja y cualquier variación en la población que favorece la adopción o la adquisición del comportamiento X se seleccionará genéticamente, de modo que tras unas cuantas generaciones, los descendientes son Xinos de nacimiento, que apenas necesitan ensayos para esta conducta; la Xtitud se ha desplazado, mediante el efecto Baldwin, a sus genomas, como un instinto). Aún otra posibilidad es que cada generación tenga que aprender los trucos desde cero, pero tal aprendizaje resulta fácil, ya que es un Buen Truco muy obvio.[11] Nadie supondría que el hecho de que los humanos siempre y en todas partes arrojan sus lanzas con el extremo puntiagudo por delante demuestra que tiene que haber un gen para el instinto de «lo puntiagudo por delante».

En cualquier caso, el problema epistemológico al que se enfrenta un recién nacido puede asemejarse al problema epistemológico de un lingüista de campo adulto ante una lengua extranjera, de la cual intenta plantear una gramática y un léxico; ahora bien, los métodos empleados por cada uno para reducir la brecha entre la ignorancia y el conocimiento son notablemente diferentes. El adulto plantea y comprueba hipótesis y busca pruebas que puedan confirmar o refutar sus suposiciones bien fundadas, refinando las generalizaciones, remarcando las excepciones, etc. El niño solamente balbucea, intentando aplacar sus apetencias y curiosidades y, así, paso a paso, se impulsa hacia la comprensión mediante un proceso inconsciente de ensayo y error a gran escala, con el que logra la misma proeza epistemológica que el adulto pero sin el beneficio de una teoría; es la misma forma en que la evolución por selección natural logra la hazaña de diseñar alas de diferentes especies de aves sin el beneficio de una teoría de la aerodinámica. Todo esto recuerda el contraste entre el diseño inteligente de arriba abajo (Gaudí, Turing y Picasso nos han servido de ejemplo) y la I+D darwiniana. Los niños adquieren su lengua materna gracias a un proceso cuasidarwiniano, con el que alcanzan unas competencias que son la base de la comprensión mediante un proceso que es competencia sin comprensión.

Podemos juzgar las primeras palabras (tanto en bebés actuales como en las profundidades de la historia del lenguaje en nuestra especie) como

especies sinantrópicas (véase el capítulo 6), que han evolucionado por selección natural (genética y cultural) para prosperar en compañía humana, con todas sus particularidades referentes a fisiología, hábitat y necesidades. Es probable que la sinantropía sea la vía hacia la domesticación que han recorrido la mayoría de nuestras especies domesticadas. Por ejemplo, tal como exponen Coppinger y Coppinger (2001), el mito de los lobos que se vuelven perros cuando intrépidos domesticadores sacan de su guarida a los cachorros de lobos salvajes es solo eso, un mito, apenas plausible. Con toda seguridad, cuando los asentamientos humanos empezaron a generar concentraciones de restos de alimentos comestibles, estos vertederos se convirtieron en atractivos para los lobos salvajes, los cuales presentaban diferencias en su capacidad de soportar la proximidad a los peligrosos humanos. Los que podían soportarla con más facilidad quedaron aislados, geográficamente y reproductivamente, de sus parientes más precavidos, que se mantenían a distancia; con el tiempo, lo que surgió fue el perro de vertedero, que no pertenecía a nadie, no era el compañero ni la mascota de nadie, sino simplemente una presencia familiar en la comunidad, como los ratones, las ratas y las ardillas. Así pues, los perros se domesticaron a ellos mismos durante muchas generaciones, hasta que sus vecinos humanos se convirtieron en sus dueños, sus guardianes, sus maestros.

De modo parecido, podemos imaginar que los niños solo se dan cuenta gradualmente de que tienen todas esas palabras que pueden empezar a usar. De hecho, empiezan a usarlas, y a aprovecharse de las ventajas de usarlas, mucho antes de darse cuenta de que lo están haciendo. Finalmente, llegan a un punto en que las palabras en su imagen manifiesta se convierten en sus propias palabras, ofrecimientos que forman parte de su equipaje, como un garrote o una lanza, más que ofrecimientos que aparecen en la naturaleza, como una piedra para lanzar o un agujero donde esconderse. Sus palabras han sido domesticadas, con el proceso que Darwin resumió en el capítulo inicial de El origen de las especies (1859) y luego desarrolló en su magistral obra La variación de los animales y las plantas bajo domesticación (1868). Este empieza con lo que Darwin denomina selección inconsciente, en que las personas, sin demasiado cuidado o de forma inadvertida, favorecen a ciertos descendientes en detrimento de otros, con lo que crean una fuerza selectiva que se irá centrando y enfocando progresivamente; al final, llegamos a la selección metódica, como la de los colombófilos, los cultivadores de rosas o los criadores de caballos o ganado, que tienen en mente objetivos concretos, rasgos que intentan seleccionar. Este es un

paso importante hacia el diseño inteligente de arriba abajo, en que los domesticadores poseen razones (buenas o malas) para lo que hacen y para el resultado que esperan. En el caso de la domesticación de las palabras, esta aparece cuando los individuos empiezan a reflexionar, a ser conscientes de su propio uso del lenguaje, evitando las palabras que consideran mediocres, ofensivas o pasadas de moda (o demasiado jergales o nuevas).

Cuando las personas comienzan a editar su producción del lenguaje (literalmente, en el caso de la escritura, o figurativamente cuando construyen frases en su mente antes de expresarlas en voz alta) están cumpliendo el requisito final de Darwin para la domesticación: controlar la reproducción de los miembros de la especie, o como mínimo de aquellos que poseen.[12] Téngase en cuenta que puede haber razones por las que las personas evitan «inconscientemente» determinadas palabras o favorecen otras, pero el resultado no es un lenguaje domesticado en su totalidad hasta que poseen razones. Las palabras sinantrópicas, igual que los organismos sinantrópicos, tienen que apañárselas solas cuando se trata de la reproducción; no reciben ninguna asistencia de sus poseedores. Una vez domesticadas, ya pueden bajar un poco la guardia: su reproducción está casi asegurada, gracias al cuidado de sus guardianes. Por ejemplo, los términos técnicos escogidos cuidadosamente dentro de un ámbito científico son un paradigma de memes totalmente domesticados, enseñados pacientemente a los jóvenes con prácticas y pruebas obligatorias. No tienen por qué ser divertidos, excitantes o teñidos del aura seductora de las cosas prohibidas, como los reproductores sinantrópicos más capaces. Las palabras domesticadas tienen el respaldo de las altas esferas y se reproducen con ayuda de los métodos que han demostrado funcionar, pero deberíamos seguir la estela de Darwin y admitir que no hay solución de continuidad entre estos casos de domesticación a gran escala y los patrones de preferencias que llevan a la selección inconsciente de palabras en contextos particulares.

Los fonemas son quizá los rasgos de diseño más importantes del lenguaje humano y es muy posible que ya estuvieran a punto cuando las palabras eran sinantrópicas, características apenas identificadas del entorno memético y no valiosas posesiones de los seres humanos. Los fonemas logran la digitalización del medio auditivo para el habla. Cada idioma hablado tiene un alfabeto auditivo finito, los fonemas a partir de los cuales se forman las palabras. La diferencia entre col, mol, sol y rol es el fonema inicial, mientras que entre mar y mal es el fonema final. Cuando

uno aprende a leer con el método fónico, aprende cómo se corresponden las formas escritas con los fonos (sonidos) que componen los fonemas (unidades de sonido distintivas) de un idioma. El sonido de la be en rombo es ligeramente diferente del sonido de la be en robo, pero se trata de diferentes fonos, no de diferentes fonemas, porque no hay dos palabras (en castellano) que difieran solamente por esta diferencia en la forma en que lo hacen robo y rombo.

Lo que es fantástico de los fonemas es que son inmunes a las diferencias de acento, tono de voz y otras variaciones habituales del habla. Hay cientos de formas diferentes de pronunciar «La mantequilla está en la alacena» (en andaluz cuchicheando, en quiteño con un susurro, en rioplatense con un quejido…) y todas son identificables sin esfuerzo por casi cualquier hablante de castellano como casos de «La mantequilla está en la alacena». Hay una variación casi continua en los fenómenos físicos que atraviesan los filtros procústeos para «adaptarse a las normas» y «tomarlo o dejarlo», convirtiendo cada situación límite en un fonema u otro. Este es el fundamento de la digitalización, que obliga a fenómenos continuos a acomodarse a fenómenos discontinuos, fenómenos de todo o nada. Los fonemas protegen las fronteras de los tipos, de modo que puedan tener casos muy diversos, lo que es un requisito para una replicación fiable. Nuestros sistemas auditivos se ajustan desde antes del nacimiento (sí, en el útero) para clasificar un amplio y variado espectro continuo de sonidos en las diferentes ejecuciones de los fonemas de nuestra lengua materna. Si esperamos hasta la adolescencia para aprender otro idioma, es muy probable que tengamos dificultad con los acentos pero, sobre todo, con la percepción espontánea de las palabras emitidas por otros, ya que tendremos unos detectores de fonemas imperfectos y poco fiables.

Sin un plan de digitalización, cuesta recordar los sonidos audibles y reproducirlos, y pueden ir a la deriva imperceptiblemente, alejándose de sus antecesores y llegando al fenómeno de la catástrofe de error, en que la selección natural no puede funcionar porque las mutaciones se acumulan con más rapidez de lo que pueden ser seleccionadas, tanto a favor como en contra y, así, se destruye la información semántica que pudieran transportar. Da igual si vale la pena recordar, reproducir o copiar un sonido audible, si no es digitalizable, sus perspectivas de ser recordado son bastante malas. Los memes audibles sinantrópicos, tanto si son mutualistas, comensales o parasitarios, persistirán y se propagarán solamente si tienen partes fonémicas: laralalá, achís, tururú, bla bla bla.

Se trata de la misma digitalización que proporciona a las computadoras su extraordinaria fiabilidad. Tal como hizo notar Turing, no hay nada en la naturaleza que sea realmente digital, siempre hay variaciones continuas; la gran jugada de diseño es construir dispositivos que traten todas las señales como si fueran digitales, descartando las particularidades de los casos individuales, en lugar de copiarlas. Así es como se pueden hacer copias prácticamente perfectas de archivos de texto o de música. Podemos comparar un copiador de CD y una fotocopiadora: si fotocopiamos una fotografía, luego copiamos la fotocopia, luego copiamos la copia de la fotocopia y así sucesivamente, la imagen se irá haciendo más y más borrosa y se acumulará todo tipo de «ruido» visual; en cambio, un archivo JPEG en un CD puede descargarse en una computadora y luego utilizarse para hacer otra copia en otro CD, incontables veces y sin pérdida de claridad o de fidelidad, porque en cada copia, cualquier pequeña variación que haya en los casos de ceros y unos será ignorada en el proceso de copia. Esto también es cierto en el lenguaje escrito, con sus alfabetos finitos que normalizan una variabilidad a todas luces infinita de formas y tamaños de letras. Esto permite una transmisión de información del lenguaje hablado (y escrito) extremadamente fiable, y en ausencia de comprensión.

FIGURA 9.3. Adaptación al castellano del CAT de Selfridge

Un famoso ejemplo escrito es el que presentamos en la figura 9.3, debido a Oliver Selfridge.******* Un lector hispanohablante lo leerá como LA LECHE, aunque el segundo símbolo y el sexto tienen la misma forma intermedia. Las palabras habladas también se ven obligadas a encorsetarse automáticamente en secuencias fonémicas, en función del idioma del oyente. Unos oyentes hispanohablantes no tendrán dificultades en reproducir con exquisita precisión «me pica el cerdo» tras oírlo una sola vez, aunque no tengan la más mínima idea de qué significa (una expresión cheli para decir «me siento mal»),******** pero no serán capaces de reproducir con precisión los sonidos que podríamos transliterar como «fnurglzhnyum djyukh pssajj»; por muy clara y alta que se articule esta secuencia de sonidos, no tiene ninguna descomposición

automática en fonemas del castellano. Gracias a este sistema de normas, incluso sinsentidos como «los fulquérrimos caslos trilaban y gilateaban en la yelma» se pueden percibir con claridad y transmitir con precisión.

Los fonemas no son solo una forma brillante de organizar estímulos auditivos para su transmisión fiable, también son una clase de ilusión benévola como la ingeniosa animación ilusoria de los iconos informáticos para arrastrar y soltar, esas pequeñas carpetas sobre las cuales podemos dejar caer los archivos, y todo el resto de elementos cada vez más familiares de cualquier escritorio de una computadora actual. Lo que realmente sucede al efectuar este arrastrar y soltar sobre el icono de una carpeta es brutalmente complejo, pero los usuarios no tienen ninguna necesidad de saber qué es lo que se cuece, de modo que diseñadores de interfaces inteligentes han simplificado los ofrecimientos, haciéndolos especialmente prominentes a los ojos humanos y añadiendo efectos de sonido para ayudar a centrar la atención. No hay nada compacto y prominente dentro de la computadora que corresponda a esa pequeña carpeta que aparece en la pantalla. Tampoco hay nada compacto y prominente que se pueda discernir entre las propiedades físicas de las diferentes ocurrencias de los fonemas que diferencian col de sol y de sal o cabo de rabo y de rato. Las diferencias entre estas palabras parecen obvias y simples, pero se trata de una ilusión generada por nuestra competencia incorporada, no por una simplicidad intrínseca de la señal. Tras décadas de investigación y desarrollo, los programas de reconocimiento de voz son casi tan competentes como un niño de cinco años para extraer los fonemas del habla informal de la vorágine acústica que llega a un oído o a un micrófono.

La digitalización de los fonemas tiene una profunda implicación: las palabras tienen un papel en la evolución cultural parecido al papel del ADN en la evolución genética, pero a diferencia de los peldaños idénticos en la doble hélice, hechos de adenina, citosina, guanina y timina, las palabras no son replicadores físicamente idénticos; solo son «idénticos» como una ilusión del usuario de la imagen manifiesta. Podríamos decir que las palabras son una especie de ADN virtual, un medio digitalizado en gran parte que solo existe en la imagen manifiesta.

Los robles y los manzanos transmiten millones de bellotas y manzanas por el mundo «con la esperanza de» dar inicio a nuevos descendientes. Cuando transmitimos millones de palabras, «la esperanza» no es para nuestros descendientes, sino para los descendientes de las palabras, como

los descendientes del virus del resfriado que enviamos a nuestros colegas cada vez que estornudamos. Cuando las palabras llegan a otro cerebro, pueden «entrar por una oreja y salir por la otra» o, en ocasiones, pueden echar raíces. Uno de mis profesores en la escuela primaria solía decirnos «¡usad una palabra tres veces y será vuestra!» para aumentar nuestro vocabulario, y no iba muy desencaminado.

En este punto, algunos filósofos fruncirán el ceño y se preocuparán porque estoy atravesando un terreno resbaladizo. ¿Existen las palabras? ¿Son parte de tu ontología? ¿Deberían serlo? Toda esta charla sobre palabras

«hechas de información» es un poco arriesgada, ¿no? ¿Es solo un montón de palabrería vacía? Algunos filósofos harán de tripas corazón e insistirán en que, en sentido estricto, las palabras no existen; no tienen masa ni energía ni composición química, no forman parte de la imagen científica, que ellos consideran que debería ser el árbitro definitivo de la ontología. Pero las palabras son unos moradores muy destacados de nuestra imagen manifiesta e incluso si la ciencia no tiene que referirse a ellas ni mencionarlas, no se puede hacer ciencia sin usarlas, por lo que quizá sí que deberíamos incluirlas en nuestra ontología. Se ciernen sobre nosotros, ocupando nuestra atención.[13]

En todo esto, el contraste entre nosotros y los chimpancés es notorio: hoy en día, miles de chimpancés han pasado toda su vida en cautividad humana y han oído casi tantas palabras como los niños humanos, pero apenas les prestan atención. Para ellos el habla humana no es muy diferente del susurro de las hojas de los árboles, por mucho que el habla contenga enormes cantidades de información semántica que les podría ser de utilidad si llegaran a caer en cuenta. Les sería muy fácil escapar del cautiverio o engañar a los investigadores si pudieran escuchar y comprender las conversaciones de sus guardianes. Se necesita un ingente programa de entrenamiento para lograr que un chimpancé preste atención a las palabras, habladas, simbolizadas o en forma de fichas de plástico. Los bebés humanos, en cambio, están ansiosos por recibir experiencias verbales desde su nacimiento. Las palabras son ofrecimientos que nuestros cerebros están diseñados (gracias a procesos evolutivos) para recoger, tal como afirmó Gibson, y permiten una gran cantidad de usos de todo tipo.

******* El ejemplo original de Selfridge, en inglés, es THE CAT. Nos permitimos adaptarlo al castellano como LA LECHE, cuya lectura será más familiar e identificable para un lector hispanohablante. (N. del t.)

******** Nuevamente, nos permitimos adaptar el ejemplo original del autor a una versión más adecuada para el lector hispanohablante. (N. del t.)

10

EL PUNTO DE VISTA DE LOS MEMES

PALABRAS Y OTROS MEMES

Pienso que un nuevo tipo de replicador ha surgido recientemente en este mismo planeta. Lo tenemos frente a nuestro rostro. Se encuentra todavía en su infancia, aún flotando torpemente en su caldo primario, pero ya está alcanzando un cambio evolutivo a una velocidad que deja al antiguo gen jadeante y muy atrás. […] El nuevo caldo es el caldo de la cultura humana. Necesitamos un nombre para el nuevo replicador, un sustantivo que conlleve la idea de una unidad de transmisión cultural, o una unidad de imitación.

R

ICHARD DAWKINS, El gen egoísta

Estoy convencido de que las comparaciones entre la evolución biológica y el cambio cultural o tecnológico humano han hecho muchísimo más mal que bien, y los ejemplos abundan de esta, que es la más común de las trampas intelectuales. […] La evolución es movida por la selección natural, la evolución cultural por un conjunto diferente de principios que yo comprendo, pero nebulosamente.

S

Las palabras existen en la imagen manifiesta, pero ¿qué son? Los perros son un tipo de mamífero o un tipo de mascota. Las palabras ¿de qué son un tipo? Son un tipo de meme, un término que acuñó Richard Dawkins en su libro El gen egoísta y que queda definido en su epígrafe. ¿Qué tipo de memes son las palabras? El tipo que se puede pronunciar. Además de palabras, existen otros elementos léxicos (Jackendoff 2002) como tiempos verbales irregulares y excepciones a las «reglas» de una lengua que también tienen que almacenarse independientemente en la memoria. Lo que se transmite de otros hablantes a un hablante principiante que recibe, por ejemplo, el pretérito perfecto simple irregular de andar, es un elemento léxico. Se trata de memes paradigmáticos, no rasgos innatos del lenguaje o de la gramática, sino rasgos opcionales asociados a la cultura que tienen que propagarse mediante comunidades lingüísticas y, a menudo, en competición con otras variantes. Otros memes, como el meme de llevar la gorra con la visera sobre el cogote, el de levantar el pulgar para indicar aprobación o el de construir un arco que tiene esta forma, no tienen pronunciación y, por tanto, no son palabras.

Y los memes, ¿de qué son un tipo? Son, a grandes rasgos, un tipo de manera de comportarse que puede copiarse, transmitirse, recordarse, enseñarse, evitarse, denunciarse, mostrarse, ridiculizarse, parodiarse, censurarse, santificarse. En el lenguaje técnico de la imagen científica no hay ninguna palabra disponible con facilidad que sintetice adecuadamente qué tipo de cosa es un meme. Forzando un poco el lenguaje coloquial de la imagen manifiesta, podríamos decir que los memes son maneras: maneras de hacer algo o de construir algo, pero no instintos (que son un tipo diferente de maneras de hacer o construir cosas). La diferencia es que los memes se transmiten perceptualmente, no genéticamente.

Los memes son información semántica, diseños que vale la pena robar o copiar, excepto cuando se trata de información incorrecta que, igual que el dinero falso, es algo que se transmite o se guarda bajo la equivocada suposición de que es valioso y útil.[1] Y como hemos visto, la desinformación es diseño que vale la pena copiar. Al fin y al cabo, el correo basura de Internet, como la propaganda, está diseñado. ¿A quién

beneficia? A veces a los anfitriones, a veces a los autores y a veces a los propios memes, que pueden no tener autor pero, al igual que los virus (que tampoco tienen autores), tienen su propia aptitud, su propia capacidad de replicación relativa.

Las palabras son el mejor ejemplo de memes. Destacan con claridad y tienen una fuerte individualidad como elementos de nuestra imagen manifiesta. Sus historias de descendencia con modificación, tanto en pronunciación como en significado, son nítidas y en muchos casos pueden remontarse miles de años. Son numerables (pensemos en el tamaño del vocabulario) y su presencia o ausencia en vectores o anfitriones humanos individuales se puede detectar mediante una simple prueba. Se puede observar su propagación y ahora, gracias a Internet, disponemos de un exquisito laboratorio con el que obtener más información. Ahora bien, como siempre pasa en un laboratorio, hay que pagar un precio, ya que restringimos los fenómenos a entornos artificiales y bien delimitados y nos arriesgamos a introducir sesgos considerables cuando seleccionamos las poblaciones en estudio (no todos los usuarios del lenguaje son usuarios de Internet). No es ninguna casualidad que la población de casos de la especie meme (el término en sí mismo), fundada por Dawkins en 1976, languideciera plácidamente hasta que Internet proporcionó un nicho ideal para su uso.

Si las palabras son los mejores memes, ¿por qué Dawkins no les dio mayor protagonismo en su presentación? De hecho, empezó su capítulo sobre evolución cultural citando la acumulación de cambios en la lengua inglesa desde Chaucer hasta el presente: «Parece ser que el lenguaje "evoluciona" por medios no genéticos y a una velocidad más rápida en órdenes de magnitud que la evolución genética» (1976). Su primera lista de ejemplos de memes era «tonadas o sones, ideas, consignas, modas en cuanto a vestimenta, formas de fabricar vasijas o de construir arcos». Las consignas son elementos léxicos formados por palabras, pero Dawkins no subrayó el papel de las palabras en la transmisión de «ideas», sin duda porque las ideas se pueden compartir y transmitir sin palabras (mostrando algo, sin describirlo ni definirlo, pongamos por caso) y Dawkins quería remarcar la extensión, en absoluto obvia, del concepto de meme más allá de las palabras, a otras clases de elementos culturales: melodías, no necesariamente canciones con letras; modas en la indumentaria tanto si tienen nombre como si no, etc. Términos como minifalda o pajarita se propagan junto con sus modelos, pero también lo

hacen rasgos a los que no se presta atención, que solo se hacen patentes, y posteriormente se nombran, después de haberse multiplicado incontables veces: llevar la gorra con la visera al revés, tejanos con desgarres o calcetines con sandalias.

La principal tecnología que ha facilitado la eclosión de la genómica es la reacción en cadena de la polimerasa (o PCR), que, a partir de cualquier muestra de ADN, genera una enorme cantidad de copias, de modo que es fácilmente detectable, distinguible y manipulable. Las repetidas y machaconas copias de los anuncios que cubren las paredes de las escaleras mecánicas del metro o que tapizan las vallas de los solares en construcción captan la mirada de los observadores, creando patrones irresistibles. Como siempre, la repetición es el ingrediente básico para crear nuevos ofrecimientos, ya sean dispuestos sincrónicamente en el espacio, como los anuncios o los fragmentos idénticos de ácidos nucleicos, o a lo largo del tiempo, como ensayar una melodía o una palabra u observar una y otra vez la torre Eiffel en París. Las copias múltiples de algo tienden a hacer que nuestra maquinaria de reconocimiento de patrones pueda producir una nueva copia, en el «reconocedor» y, así, se puede propagar un meme.

En la figura 7.4 (p. 138), las palabras están representadas con un valor alto de reproducción frente a crecimiento, un valor también alto de cultura frente a genética y un valor bajo de complejidad. Quizá si Dawkins hubiera desarrollado el tema de las palabras como memes paradigmáticos, hubiera frustrado o suavizado parte de la resistencia que encontró su propuesta por parte de otros investigadores de la cultura: historiadores, teóricos de la literatura, filósofos, lingüistas, antropólogos y sociólogos, por ejemplo. Ninguno de ellos había conjeturado nunca que las palabras fueran creaciones diseñadas deliberadamente (excepto en el caso de un puñado de palabras acuñadas). Como nunca ha habido ningún mito de palabras como artefactos culturales diseñados inteligentemente, no habrían tenido ningún interés particular en proteger un mito así y podrían haber sido más abiertos a la idea de que había bastantes más elementos de la cultura parecidos a las palabras de lo que habían creído. En cualquier caso, podemos afirmarlo ahora y pasar a recalcar que una vez que las palabras quedan aseguradas como el medio dominante de innovación y transmisión cultural, empiezan a transformar el proceso evolutivo y dan lugar a nuevas variedades de I+D mucho más cercanas a la idea tradicional y mítica del diseño inteligente.

¿QUÉ TIENEN DE BUENO LOS MEMES?

El precio que pagamos por nuestro promiscuo deseo de información adaptativa es acoger algunas variantes culturales de una patología asombrosa.

R

ICHERSON Y BOYD, Not by Genes Alone

Entre los científicos que estudian la evolución cultural, se ha producido una cierta prohibición contra el uso del término meme de Dawkins (1976) para referirse a las maneras de hacer y construir cosas que se propagan por las culturas, pero todos los teóricos que recuerdo emplean ese mismo concepto bajo un término u otro: ideas, prácticas, métodos, creencias, tradiciones, rituales, términos, etc. Son todos ellos objetos informacionales que se propagan por los seres humanos más o menos del mismo modo que los gérmenes o los virus. La información es con lo que trafica la evolución cultural, como la evolución genética y, tal como expresan Richerson y Boyd (2005):

necesitamos algún acuerdo conveniente acerca de cómo denominar a la información almacenada en los cerebros de las personas. No es un problema trivial, porque los psicólogos tienen desacuerdos profundos sobre la naturaleza de la cognición y del aprendizaje social. (p. 63)

Cuando hablamos sobre información que ha evolucionado culturalmente, bien pocas veces estamos hablando de bits de información (al modo de Shannon), y sería bueno disponer de un término general para un trozo o fragmento relevante de información, un término que pudiera acompañar y contrastar el término gen. La palabra meme ya ha logrado establecerse

en la lengua inglesa, y aparece en la última edición del Oxford English Dictionary definido como «un elemento de la cultura que puede considerarse que se transmite por medios no genéticos»,******** de modo que podemos aceptarla como el término genérico para cualquier manera basada en la cultura. Los que tengan remordimientos por usar un término cuya identidad aún está bajo asedio, pueden pensar que todavía abundan las controversias acerca de la definición de su homólogo, gen, un término que casi nadie recomendaría abandonar. Así, rebelándome contra la tradición de evitar el término, seguiré denominando memes a la mayoría de elementos culturales y apoyaré esta terminología a medida que avancemos, pues creo que las contribuciones que hace el concepto de Dawkins a nuestra comprensión de la evolución cultural compensan con creces cualquier posible connotación desafortunada que haya podido adquirir.

¿Por qué el concepto de meme se granjeó esta mala reputación? En parte fue gracias a algunas exageraciones planteadas por algunos aspirantes a memetistas que no hicieron bien sus deberes y en parte a algunos teóricos que ya habían descifrado algunos aspectos de la evolución cultural y no querían conceder a Dawkins un mérito indebido adoptando su terminología al hablar de sus propios trabajos teóricos. También hubo algunas críticas esmeradas de diversos aspectos de la descripción de los memes de Dawkins, en las que nos detendremos a su debido tiempo. Pero en mi opinión, el factor más influyente ha sido una frenética campaña de críticas descabelladas, una especie de reacción alérgica por parte de académicos de las ciencias humanas y sociales que se inquietaron al ver que esa espantosa biología estaba invadiendo sus recintos sagrados.[2] En el capítulo 11 responderé a las objeciones habituales procedentes de este campo, tras explicar algunas de las ventajas de adoptar el punto de vista de los memes y también tras responder a las objeciones realmente serias que se plantean.

Entre los avances que los memes de Dawkins nos ofrecen para el estudio de la evolución cultural, cabe destacar tres aspectos:

Competencia sin comprensión. El carácter diseñado de algunos elementos culturales no se puede atribuir a ningún autor ni grupo de autores, a ningún arquitecto, a ningún diseñador inteligente; la

innegable inventiva o aptitud de la disposición de las partes se debe por completo a la selección natural: la replicación diferencial de simbiontes informacionales cuyos anfitriones pueden ser tan ciegos acerca de su inventiva como las mariposas acerca de los ocelos de sus alas (los seres humanos aprecian a menudo las virtudes de las ideas que propagan con entusiasmo, pero como veremos, esto se trata de un rasgo reciente y opcional de buena parte de la evolución cultural; la comprensión y aprobación humanas no son necesarias ni suficientes para que un meme quede fijado en una cultura).

La aptitud de los memes. Los memes tienen su propia aptitud reproductiva, como los virus. Tal como expresa Dawkins, «[l]o que no hemos considerado previamente es que una característica cultural puede haber evolucionado de la manera que lo ha hecho simplemente porque es ventajoso para ella misma» (1976 [1989]). Tanto si un meme mejora la aptitud reproductiva de sus anfitriones como si no (los seres humanos que adoptan y usan los diseños que transporta), la selección natural de los memes está haciendo el trabajo. Los memes pueden propagarse hasta fijarse en una sola década y extinguirse en un intervalo parecido, un proceso evolutivo demasiado rápido para que sea discernible en cualquier aumento o disminución de descendencia humana (incluso si todos los memes fueran simbiontes mutualistas, adaptaciones importables que mejoran la aptitud reproductiva de sus anfitriones, la velocidad de su propagación no se podría explicar por este minúsculo efecto).

Los memes son objetos informacionales. Son «instrucciones» para maneras de hacer cosas, que se pueden transmitir, almacenar y mutar sin ejecutarse o expresarse (como los genes recesivos, que viajan sigilosamente en el genoma). Se cree que Marco Polo trajo a Europa el meme de la pasta desde China; no tuvo que convertirse en un chef de la pasta para hacerlo, le bastó con dispersar copias del meme en entornos en que otros seres humanos podían quedar infectados por él y expresarlo en sus maneras de comportarse.

Las teorías tradicionales del cambio cultural en las humanidades y las ciencias sociales se han esforzado sin estas ideas. El viejo funcionalismo durkheimiano, por ejemplo, descubrió muchos propósitos y funciones

muy plausibles de diversas convenciones sociales (tabús, prácticas, tradiciones, distinciones sociales, etc.), pero nunca pudo explicar cómo aparecieron estas eficaces convenciones. ¿Fue a causa de la inteligencia de los reyes, caciques o sacerdotes? ¿Un regalo de Dios? ¿Pura suerte? ¿O fue quizá gracias a alguna misteriosa «mente grupal» o a una «genialidad popular» compartida que entendía estas características y por ello condujo a su adopción? La teleología no sale gratis, y los durkheimianos no tenían ninguna forma defendible de pagar su precio. Pero es que no es necesario postular mecanismos mágicos; la selección natural de los memes puede hacer el trabajo de diseño sin ninguna ayuda de la comprensión humana, divina o grupal. Tal como remarca David Sloan Wilson (2002), el funcionalismo murió por falta de un mecanismo creíble. La teoría de la evolución de los memes por selección natural puede aportar el mecanismo necesario, tal como la tectónica de placas aportó el mecanismo necesario para que la hipótesis de la deriva continental fuera tomada en serio.[3]

Otra dificultad con las teorías tradicionales de la cultura es que son propensas a establecer categorías «psicológicas» para sus entidades informacionales: en concreto, ideas o creencias. Pero, aunque es evidente que las ideas y las creencias (por más que queramos caracterizarlas como estados o episodios mentales o psicológicos) tienen un papel destacado en la cultura humana, no toda la transmisión y evolución cultural depende de algo como una «aceptación consciente». Un grupo puede variar imperceptiblemente su pronunciación de una palabra, sin darse cuenta del cambio. Los significados de las palabras también pueden evolucionar mediante procesos alejados de la comprensión de los usuarios de las palabras, gracias a la replicación diferencial. Hoy en día, cuando alguien afirma con entusiasmo «¡fue una conferencia increíble, simplemente alucinante!» no quiere decir que nadie la creyera y que provocara alucinaciones en los oyentes. Nadie decidió un buen día ajustar estos significados ni tampoco aprobar el cambio; se dio por una variación en la población de casos producidos en la cultura (véase Richard, próxima publicación, para más detalles sobre este tema).

Cuando se adopta la perspectiva psicológica tradicional de ideas y creencias, el hecho de que los cambios en los rasgos culturales se pueden propagar sin que nadie se dé cuenta no tiene una explicación fácil (y, por tanto, es probable pasarlo por alto). El punto de vista de los memes proporciona una valiosa corrección a este descuido, pero es mucho más

relevante la manera en que los memes pueden ofrecer una visión alternativa de cómo la información transmitida por la cultura se instala en los cerebros sin ser comprendida. El problema con el punto de vista habitual es que se basa acríticamente en la suposición de que la racionalidad está incorporada en el enfoque intencional: la presunción predeterminada de la psicología popular es que las personas, e incluso los animales «superiores», comprenderán todo aquello que se les plantee ante sus ojos. Ideas, creencias y conceptos nuevos son ideas, creencias y conceptos comprendidos, casi por definición. Tener una idea es saber qué idea tienes, imaginarla y diferenciarla con claridad, tal como dijo Descartes. Apenas nos damos cuenta de que siempre que intentamos identificar una idea para comentarla o discutirla, la identificamos por su contenido (es decir, no como «la primera idea que ha mencionado Tomás» o «la idea que guardo justo al lado de mi idea del fútbol»). Esto es una aplicación omnipresente de la noción predarwiniana de que la comprensión es el origen de la competencia, y es importante que invirtamos esta idea, de modo que podamos esbozar, por primera vez, una descripción de cómo competencias sin comprensión pueden generar comprensión, paso a paso (tema que nos ocupará el capítulo 13).

Los teóricos tradicionales de la cultura suelen sobreestimar la contribución de artífices individuales, suponiendo que deben de haber tenido más comprensión de los diseños que inventan, transmiten y mejoran de la que sería estrictamente necesaria. Pero la comprensión no es necesaria para lograr mejoras de diseño. En un estudio de la evolución de las canoas polinesias, Rogers y Ehrlich (2008) citan un fragmento de Alain ([1908] 1956), un filósofo francés (que escribe acerca de barcas de pesca de la Bretaña, no de canoas polinesias):[4]

Cada barca es una copia de otra barca […] Razonemos como sigue, a la manera de Darwin. Está claro que una barca mal construida acabará en el fondo del mar tras uno o dos viajes y no se copiará […] Diríamos, con todo rigor, que es el mar quien moldea las barcas, escogiendo aquellas que funcionan y destruyendo las otras.

Los constructores de barcas pueden tener teorías, buenas o malas, sobre qué podría constituir una mejora en el arte que sus padres y abuelos les

han enseñado; cuando incorporan estas revisiones en sus creaciones, a veces la acertarán y a veces no. Otras innovaciones pueden no ser más que errores de copia, como las mutaciones genéticas; la probabilidad de que estos cambios fortuitos sean una mejora es muy baja, pero muy de vez en cuando alguno de ellos hará progresar el arte local de la construcción de barcas. Las barcas que vuelven se copian; el resto cae en el olvido, en un ejemplo claro de selección natural. Las barcas copiadas pueden acumular elementos innecesarios pero inocuos (como decoraciones) además de sus adaptaciones genuinamente funcionales; posteriormente, estos elementos meramente decorativos pueden adquirir una función, o no. Así, el punto de vista memético puede amoldarse a mejoras tanto funcionales como meramente tradicionales de los artefactos, rituales y otras prácticas humanas, y también puede explicar nuestra incapacidad de trazar una frontera clara que distinga las dos categorías sin recurrir a la inescrutabilidad de la «genialidad popular» que las creó. No se requiere ningún tipo de comprensión, aunque es posible que esta acelere los procesos de I+D con más frecuencia que los retarde.[5]

Otra carencia de buena parte del pensamiento tradicional sobre la cultura es la tendencia a concentrarse en las cosas buenas e ignorar las inútiles. Adoptar un modelo económico de la cultura es adecuado como primera aproximación para sus elementos claramente útiles (o simplemente benignos): tecnología, ciencia, arte, arquitectura, literatura… su «alta cultura», en resumidas cuentas. Sin duda esto son tesoros de la sociedad y tiene sentido económico (es racional) conservarlos y cuidarlos, gastar lo que sea necesario para mantenerlos y asegurarse de que pasen a la siguiente generación como un espléndido legado, aunque pueda llevar años de rigurosa formación y arduo entrenamiento lograr que continúen «sanos y salvos» siglo tras siglo.[6] Pero ¿qué hay de los elementos malignos, inútiles, molestos, reductores de aptitud de una cultura? ¿Qué sucede con los elementos que no son apreciados por nadie, pero demasiado costosos para erradicarlos, como el virus del resfriado o la malaria? El punto de vista de los memes abarca ambos extremos y toda la zona intermedia, gracias a la distinción tripartita de la biología entre simbiontes mutualistas, comensales y parasitarios (véase el capítulo 9 y la nota 7, p. 384).

En el capítulo 7 ya hice hincapié en que muchas personas sienten aversión por la idea de que los memes sean como virus, porque creen, sin

razón, que los virus siempre son perjudiciales para nosotros. En absoluto; de hecho, solo una minoría muy reducida de los billones de virus que habitan en cada uno de nosotros son tóxicos de un modo u otro. Aún no sabemos con seguridad si hay algunos virus que sean buenos para nosotros o incluso mutualistas forzados como la microflora intestinal, sin la cual moriríamos.

¿Necesitamos determinados virus para prosperar? Quizá. En cualquier caso, sí que necesitamos gran cantidad de nuestros memes. Robinson Crusoe de Daniel Defoe (1719) es una auténtica enciclopedia de los memes que uno necesita para salir adelante y es probable que un adulto medio del siglo XVIII estuviera mejor equipado con memes para una vida autosuficiente que un adulto medio del siglo XXI del mundo desarrollado (lo que resulta un interesante vuelco histórico). ¿Cuántos de nosotros sabríamos cómo recoger semillas para la siembra del año próximo, encender fuego sin un encendedor o talar un árbol?[7]

¿De verdad que los memes parasitarios pueden prosperar? La respuesta afirmativa a esta pregunta fue uno de los puntos más llamativos de la introducción de Dawkins a los memes y es posible que sea la causa de buena parte de la antipatía que suscitan. Para muchos lectores, Dawkins parecía afirmar que toda la cultura es una horrible enfermedad, una infección del cerebro que provoca una actividad autodestructora; pero no es eso lo que dijo. Un caso notorio de esta equivocada suposición se halla en David Sloan Wilson (2002), que sin duda ve la necesidad de un fundamento evolutivo de la cultura:

El éxito de algunos [experimentos sociales] y el fracaso de otros sobre la base de sus propiedades respectivas constituye un proceso de variación ciega y de retención selectiva que actúa durante el curso de la historia humana reciente, además del pasado evolutivo distante. (p. 122)

Sin embargo, rechaza tajantemente los memes, porque cree que son «parasitarios» (p. 53), y así nunca considera la posibilidad de que los memes puedan ofrecer una descripción evolutiva de algunos fenómenos culturales (y de las religiones, en especial) más efectiva y comprobable que su propia hipótesis de la «selección grupal multinivel». De hecho, no

se da cuenta de que, en ocasiones, también alude a los memes: por ejemplo, encuentra útil caracterizar los catecismos «como "genomas culturales" que contienen, en una forma fácilmente replicable, la información necesaria para desarrollar una comunidad adaptativa» (p. 93). En otras ocasiones (p. ej. pp. 118-189, 141, 144) no repara en que las alternativas meméticas a su hipótesis (como mi exposición en Romper el hechizo) no solo están disponibles, sino que son más plausibles.[8]

Muchos memes, quizá la mayoría, son mutualistas, mejoras protésicas de nuestras adaptaciones existentes que mejoran la aptitud (como nuestros sistemas perceptuales, nuestras memorias, nuestras capacidades locomotoras y manipuladoras). Si no hubiera memes mutualistas desde el principio, es difícil ver cómo se hubiera podido originar la cultura (tema sobre el que volveremos más adelante). Pero una vez que ya se ha diseñado e instalado la infraestructura de la cultura (mediante una interacción entre la evolución genética y la cultural, como veremos), la aparición de memes parasitarios que explotan esa infraestructura es una posibilidad más o menos garantizada. Richerson y Boyd (2005, p. 156) denominan a tales memes «variantes culturales solitarias». Para poner un ejemplo bien reciente de este fenómeno, Internet es un artefacto complejo y costoso, diseñado inteligentemente y creado para un propósito práctico o vital: el Internet de hoy es el descendiente directo del ARPAnet, financiado por el ARPA (actualmente DARPA, la Defense Advanced Research Project Agency de Estados Unidos), creada por el Pentágono en 1958 como respuesta a la puesta en órbita del Sputnik por parte de la Unión Soviética y cuyo objetivo era facilitar la I+D de tecnologías militares. Actualmente, el correo basura, la pornografía (y las fotos de gatitos y demás memes de Internet) empequeñecen a cualquier intento de compartir diseños de alta tecnología de laboratorios respaldados por el Pentágono.

Antes de dar por sentado el conveniente estereotipo de toda esta basura como ejemplos de memes que no mejoran la aptitud humana, deberíamos recordar qué significa aptitud en el contexto de la biología evolutiva: no se trata de la salud, ni de la felicidad, ni de la inteligencia, ni de la comodidad, ni de la seguridad; se trata de la capacidad procreadora. ¿Qué memes hacen que sus portadores tengan más nietos que la media? Muy pocos, todo hay que decirlo. Y muchos de nuestros memes más apreciados claramente reducen la aptitud en el sentido biológico. Tener

educación universitaria, por ejemplo, tiene un efecto de reducción de la aptitud tan y tan claro que si, pongamos por caso, comer coles tuviera un efecto parecido, seguramente debería llevar una advertencia sanitaria: «¡CUIDADO! Comer col puede reducir el número de nietos por debajo de la media».[9] Cuando pregunto a mis estudiantes si este hecho sorprendente les asusta, dicen que no, y yo les creo; piensan que hay cosas más importantes en la vida que tener más éxito reproductivo que sus compañeros, una actitud que, si no es universal, es la dominante en toda la humanidad. Este hecho nos separa de todas las otras especies; somos la única que ha conseguido colocarse en una perspectiva que destierra la aptitud genética como propósito máximo, como summum bonum de la vida.[10] La búsqueda de descendencia es lo que hace que las aves construyan nidos y los castores, presas; lo que hace que las ballenas migren docenas de miles de kilómetros y que algunas arañas macho se lancen a su muerte al abrazo de sus parejas. Ningún salmón esforzándose en su camino aguas arriba puede reconsiderar la situación y contemplar una vida aprendiendo a tocar el violín; nosotros, sí.

Somos la única especie que ha descubierto otras cosas por las que vivir (y por las que matar): la libertad, la democracia, la verdad, el comunismo, el catolicismo, el islam y muchos otros complejos meméticos (memes hechos de memes). Como todos los seres vivos, nacemos con fuertes sesgos a favor de permanecer con vida y procrear, lo que se manifiesta en los deseos irresistibles de nuestra «naturaleza animal» y muchos otros hábitos y propensiones más sutiles. Pero somos una especie influenciable, no solo capaz de aprender, como las aves o los monos, o de ser entrenada, como los perros o los caballos (y los animales de laboratorio en regímenes muy estrictos), sino que también podemos ser impelidos por razones, razones que se nos representan, no motivaciones flotantes. Ya hemos visto, con las termitas que construyen castillos y los antílopes saltarines, algunos ejemplos en que hay razones que no son las razones del organismo. Cuando los animales hacen cosas por alguna razón, su comprensión de tales razones (de por qué están haciendo lo que están haciendo) o es inexistente o es muy limitada, tal como muestran gran cantidad de experimentos que demuestran su incapacidad de generalizar, de usar lo que parecen entender en versiones novedosas o contextos más amplios.

En cambio, nosotros no hacemos cosas solo por razones; a menudo tenemos razones para lo que hacemos, en el sentido de que las hemos

articulado para nosotros mismos y las hemos respaldado tras una debida reflexión. Nuestra comprensión de las razones que podemos exponer por actuar como lo hacemos puede ser imperfecta, embarullada o incluso autoengañosa, pero el hecho de que podamos poseer esas razones (con buen tino o mal tino) nos hace susceptibles de poder discutirlas y de poder transmitirlas a otros. Cuando cambiamos de opinión con un poco de persuasión (a veces nuestra propia persuasión), siempre hay la posibilidad de que no «nos tomemos muy en serio» nuestra aceptación y conformidad de la revisión de las razones, a pesar de lo que podamos decir impulsivamente. La lección puede no tener el efecto a largo plazo que esperaba nuestro interlocutor o maestro en nuestras convicciones y actitudes. Es por ello por lo que practicamos, revisamos, reelaboramos, comprobamos y repasamos nuestras razones, con la práctica del ofrecimiento y la aceptación (o rechazo) de razones subrayada por Wilfrid Sellars (1962) (véase el capítulo 3 acerca del «espacio de razones»).

Aristóteles caracterizaba nuestra especie como el animal racional y Descartes y muchos otros, hasta el día de hoy, han atribuido nuestro talento como razonadores a una especial res cogitans, o 'cosa pensante', que Dios habría implantado en nuestro cerebro.[11] Si nuestra racionalidad no procede de Dios, ¿cómo pudo evolucionar? Mercier y Sperber (2011) sostienen que la capacidad de razonar de los seres humanos individuales, de expresar y evaluar argumentos lógicos, surge a partir de la práctica social de la persuasión y, de hecho, presenta restos fósiles de su origen. El clarísimo sesgo de confirmación es nuestra tendencia a destacar las pruebas positivas de nuestras creencias y teorías actuales mientras ignoramos todas las pruebas negativas. Este y muchos otros tipos de error en el razonamiento humano sugieren que nuestras habilidades fueron pulidas para tomar partido, para persuadir a otros en las discusiones, no necesariamente para obtener la buena solución. Nuestro talento subyacente «favorece decisiones de fácil justificación, pero no necesariamente mejores» (p. 57). El proceso evolutivo de I+D que podría diseñar tal habilidad tiene que depender, por fuerza, de alguna habilidad previa del uso del lenguaje, si bien incipiente, de modo que podemos considerarlo un proceso coevolutivo, en parte evolución cultural y en parte evolución genética, liderado por la evolución cultural de memes pronunciables, las palabras.

******** En castellano, el término meme también está bien establecido, aunque en el momento de la publicación de este libro, la RAE todavía no lo ha admitido en su Diccionario. (N. del t.)

11

PROBLEMAS CON LOS MEMES: OBJECIONES Y RÉPLICAS

¡LOS MEMES NO EXISTEN!

En varias ocasiones, cuando me he encontrado con hostilidad hacia los memes, he recibido burlas altivas: «¿Memes? ¿Memes? ¿Me puedes demostrar que existen?». Las afirmaciones de no existencia casi siempre resultan resbaladizas, especialmente cuando los científicos se ponen filosóficos o los filósofos se ponen científicos. Creo que podemos estar de acuerdo (¿verdad?) en que las sirenas, los duendes, el flogisto y el élan vital no existen, pero en algunos ambientes se cuecen controversias sobre los genes, las cuerdas (de la teoría de cuerdas), el registro formal del lenguaje (en contraposición a los idiolectos), los números, los colores, el libre albedrío, los qualia o incluso los sueños. A veces, las afirmaciones negativas se basan en una doctrina de la realidad muy austera, según la cual en realidad no existe nada excepto átomos y vacío (o partículas subatómicas y campos físicos, o solo una partícula subatómica que va y vuelve entre el principio y el final del tiempo, hilando un tejido con ella misma —una propuesta del físico John Archibald Wheeler—). Filósofos y científicos por igual se han sentido siempre atraídos por estos puntos de vista minimalistas. Tal como hizo notar un estudiante mío:

Parménides es el filósofo que dijo «solo existe una cosa, y no soy yo».

En ocasiones, las afirmaciones negativas generalizadas engloban toda la imagen manifiesta: los elementos de la ontología oficial de la imagen científica existen, pero no los objetos sólidos, los colores, las puestas de sol, los arco iris, el amor, el odio, los dólares, los goles, los abogados, las canciones, las palabras, etc. Son ilusiones útiles, quizá, como la ilusión del usuario de los iconos del escritorio de una computadora. Los patrones de píxeles coloreados en la pantalla son reales, pero representan entidades tan ficticias como el pato Donald o Mickey Mouse. De forma parecida, algunos pueden afirmar que la imagen manifiesta tiene visos de realidad como una colección de imágenes (acaso la película en la que vivimos), pero que sería un error pensar en las «cosas» con las que interactuamos, manipulamos y nos enamoramos como realidad.

Supongo que es una postura defendible. De hecho, es una versión de lo que he dicho acerca de la imagen manifiesta de cada especie: una ilusión de usuario diseñada con brillantez por la evolución para ajustarse a las necesidades de sus usuarios. Mi versión solo difiere por el hecho de que está dispuesta y ansiosa por admitir estas ontologías como formas de desguazar la realidad; no simples ficciones, sino versiones diferentes de lo que realmente existe: patrones reales. La forma más impactante de decirlo («vivimos en un mundo de ficción, en un sueño, en una irrealidad») tiende a provocar confusiones porque sugiere que somos algo parecido a víctimas, embaucadas por alguna fuerza maligna. Algunos teóricos niegan la existencia de los dólares, por ejemplo, pero no tienen dificultad alguna para aceptar el amor, el odio y los arco iris; sin embargo, la realidad del amor, el odio y los arco iris nunca se menciona, porque su existencia les resulta «obvia» (¡pero no así la existencia de los dólares!).

Dejando de lado estas doctrinas metafísicas radicales, cuando se niega la existencia de X, lo que se suele implicar es que alguna determinada teoría de X es una mala teoría. Cuando la teoría es una «teoría popular» extraída de la imagen manifiesta, esto suele ser innegable. Por ejemplo, la teoría popular de los colores está repleta de errores y malentendidos, pero ¿significa esto que los colores no existen? ¿Estamos tan engañados acerca de los colores como lo estaríamos acerca de las sirenas si creyéramos en ellas? Algunos así lo afirman; otros no están tan convencidos de seguir esta senda. ¿Deberíamos demandar a Sony por publicidad engañosa al anunciar televisores en color? Una postura

intermedia afirma que los colores no existen realmente, pero que actuar como si lo hicieran es una «ficción útil».

Otra opción insiste en que los colores sí que existen, pero no son lo que la sabiduría popular cree que son. Es el camino que he seguido durante toda mi carrera, explicando que no son reales solamente los colores, también lo es la conciencia, el libre albedrío y los dólares. Este último ejemplo es interesante. Seguramente uno de los apoyos subliminales más potentes para la creencia en la realidad de los dólares procede de la clara existencia de monedas y billetes de dólares, moneda legal que puedes ver, pesar, manipular y transportar. Los bitcoins, en cambio, parecen bastante más ilusorios a ojos de muchos, pero si nos paramos a pensarlo un poco, nos daremos cuenta de que los objetos físicos que son las monedas de dólar contantes y sonantes son una especie de soporte ontológico, como unas muletas que puedes descartar una vez has aprendido a caminar y, así, solo quedarán los hábitos y expectativas mutuas que, en principio, pueden respaldar tanto a los bitcoins como a los dólares. Bien entrado ya el siglo XX, hubo acalorados debates políticos sobre la necesidad de mantener tales muletas (como el estándar de oro y los certificados de plata), pero los niños del futuro podrán crecer solo con tarjetas de crédito en sus carteras, y no por eso estarán peor. Los dólares son reales; lo que sucede es que no son lo que podríamos pensar que son.

La conciencia existe, pero tampoco es lo que la sabiduría popular cree que es; y lo mismo pasa con el libre albedrío. Sin embargo, he aprendido que algunas personas están tan convencidas de que saben lo que la conciencia o el libre albedrío tendría que ser para que fueran reales, que rechazan mis afirmaciones como hipócritas: sostienen que intento endosar un sustituto barato por la cosa de verdad. Por ejemplo:

Por supuesto, aquí el problema radica en la afirmación de que la conciencia es «idéntica» a estados físicos del cerebro. Cuanto más intentan Dennett et al. explicarme lo que quieren decir con esto, más convencido estoy de que lo que quieren decir es que la conciencia no existe. (Wright 2000, cap. 21)

[Dennett] no establece el tipo de libre albedrío y responsabilidad moral absolutas en que la mayoría de la gente quiere creer y cree. No se puede hacer, y lo sabe. (Strawson 2003)

Según mis teorías, la conciencia no es un fenómeno afísico y el libre albedrío no es un fenómeno aislado de la causalidad, y en opinión de Wright y Strawson (y otros), yo debería tener la valentía de admitir que ni la conciencia ni el libre albedrío existen realmente (quizá podría amortiguar el golpe siendo un «ficcionalista», insistiendo en que no, no existen realmente pero que resulta muy útil actuar como si existieran). No veo por qué mis críticos creen que su comprensión de lo que existe realmente es superior a la mía, de modo que objeto sus afirmaciones.

En cualquier caso, lo que debería decir sobre los memes es más sencillo. La teoría que defiendo afirma, literalmente, sin ninguna metáfora ni advertencia, que las palabras son memes que se pueden pronunciar. Los memes existen porque las palabras son memes, y las palabras existen, al igual que otras maneras de hacer cosas que se transmiten no genéticamente. Si el lector es de los que niega la existencia de las palabras, tendrá que buscar una refutación en otro lado, pues estoy conforme con incluir en mi ontología, sin más discusión, todo el mobiliario de la imagen manifiesta (palabras incluidas, por supuesto). Ahora, pasemos a considerar las objeciones relevantes a la teoría de los memes que yo defiendo.

LOS MEMES SON DISCRETOS Y TRANSMITIDOS CON FIDELIDAD. EL CAMBIO CULTURAL, NO

Richerson y Boyd (2005), aunque simpatizan con muchas de las afirmaciones de Dawkins, se resisten a denominar meme a cualquier elemento de información cultural a causa de sus connotaciones de «entidad pseudogenética discreta y fielmente transmitida, y tenemos buenas razones para creer que buena parte de la información transmitida culturalmente no es ni discreta ni fielmente transmitida» (p. 63). Ya

hemos visto que las palabras son entidades bastante «discretas y fielmente transmitidas» y que son más o menos «pseudogenéticas» o, aún mejor, que los genes son más o menos «pseudovérbicos», tal como vimos en el capítulo 8, en la analogía de Dawkins (2004) con las subrutinas de un Mac:

En cierto sentido, sin embargo, la analogía del diccionario es engañosa. Las palabras son más cortas que los genes, razón por la cual algunos autores han equiparado cada gen a una frase. Pero la analogía de las frases tampoco es buena, aunque por otro motivo: los libros no se escriben permutando un repertorio fijo de frases. La mayoría de las frases son únicas. Los genes, al igual que las palabras, pero a diferencia de las frases, se usan una y otra vez en contextos diferentes. En el contexto genético, una analogía mejor que la de las palabras o las frases es la de las subrutinas toolbox («caja de herramientas») de un ordenador. (p. 260)

Así, las palabras son como los genes en el sentido de ser estructuras informacionales (como las subrutinas) que determinan la manera en que se deben hacer ciertas cosas. Richerson y Boyd tienen razón en que «buena parte de la información transmitida culturalmente no es ni discreta ni fielmente transmitida», tal como veremos, de modo que quizá su objeción, aplicada a este punto de vista, es que los memes son solo una parte pequeña, acaso secundaria, de la información que la evolución cultural descubre y perfecciona. Supongamos que las palabras fueran los únicos memes, o casi; en tal caso, la memética o punto de vista de los memes apenas podría considerarse una teoría general de la evolución cultural. De modo parecido, Dan Sperber (2000), otro crítico importante de los memes, los define como «replicadores culturales propagados gracias a la imitación» y admite que, si bien existen unas pocas entidades de tal tipo (como las cadenas de mensajes, pero ignora las palabras como ejemplo), representan una fracción minúscula de la transmisión cultural:

Para que la memética fuera un programa de investigación razonable, debería darse el caso que la copia y el éxito diferencial en provocar una

multiplicación de las copias tuviera un papel principal abrumador en la conformación de todos los contenidos de la cultura (o como mínimo su mayoría). (p. 172)

Creo que se trata de una exigencia demasiado fuerte, y diría que Sperber podría estar de acuerdo, dada su vacilación entre el «papel abrumador» y el «como mínimo su mayoría» de los contenidos de la cultura. Sea como fuere, ¿son las palabras los únicos buenos memes? Incluso si así fuera, todavía podríamos encontrar que depender de estos memes para una cultura acumulativa, como es la nuestra, otorga al punto de vista de los memes un lugar dominante en las teorías de la evolución cultural, por más pequeña que fuera la parte de información transmitida en forma de palabras. Pero, de hecho, en la cultura humana hay más replicadores de alta fidelidad que los que estos críticos reconocen, y todos dependen, de un modo u otro, de cierta variedad de digitalización, como los fonemas, que si recordáis, crean casos que son subjetivamente similares (identificables al instante, categorizables sin discusión) a pesar de sus enormes diferencias físicas, gracias a las características con las que han evolucionado los sistemas auditivo y vocal humanos. Los casos descendientes de gato no son réplicas físicas de sus predecesores, sino que podríamos decir que son réplicas virtuales, que dependen de un sistema finito de normas de acuerdo con las cuales los hablantes corrigen inconscientemente sus percepciones y sus expresiones; y es precisamente esto (no la replicación física) lo necesario para la transmisión de información con una alta fidelidad.

Vemos el mismo tipo de abstracción, y la misma corrección a la norma, en la música (aquí me referiré a la música tonal occidental, cuya teoría conozco mejor y también será la más familiar para los posibles lectores de este libro). Nadie inventó la música tonal (do, re, mi, fa, sol, la, si), pero muchos músicos y teóricos musicales han contribuido a codificarla. Escogieron las sílabas a cantar para cada tono y perfeccionaron el sistema de notación musical; una hermosa mezcla de evolución cultural darwiniana y de diseño inteligente a lo largo de cientos de años, empezando en el siglo XI. La música tonal es un buen ejemplo de alfabeto digitalizado que permite corregir hacia la norma (estás desafinando un poco… ¡corrígelo!). Muchas innovaciones musicales se basan en efectuar disminuciones, bends, glissandi, coloraturas, etc., pero

detrás siempre están los tonos canónicos. ¿Puedes cantar o tararear la melodía Greensleeves? Si es así, estás creando un nuevo caso de esa melodía, uno más de su larga estirpe, que no está en ninguna tonalidad particular ni en ningún tempo concreto. Greensleeves interpretada vertiginosamente en do menor por un saxofonista de bebop es un caso del mismo tipo que Greensleeves interpretada melancólicamente en mi menor por un guitarrista. Si la antigua melodía no fuera del dominio público, sus derechos de autor valdrían para cualquier tonalidad y para cualquier tempo con cualquier voz o cualquier instrumento: una colección de sucesos audibles con poca semblanza física pero con una identidad detectable al instante dentro del mundo de las melodías, una parte importante de nuestra imagen manifiesta. Además, esta melodía puede escribirse fácilmente con notación musical, que la conserva del mismo modo que la escritura conserva las antiguas sagas de transmisión oral. Aun así, independientemente de la notación musical, los sistemas no escritos de escalas (como los sistemas del lenguaje antes de la invención de la escritura) ya eran una digitalización suficiente para conservar melodías y armonías simples «de oído».

En la literatura se dan niveles aún más altos de abstracción que proporcionan una «casificación» de alta fidelidad de los tipos. Dos ediciones diferentes de Moby Dick (quizá en tipos de letra diferentes) cuentan como copias de la misma novela, pero también sus traducciones a otros idiomas (en caso contrario, ¿podrías decir que has leído la Ilíada de Homero o Guerra y Paz de Tolstói?). La traducción es un proceso que requiere comprensión (como mínimo hasta que se elaboró Google Translate) pero eso no impide que la información fiable salte la barrera del idioma y, así, se pueda producir una diseminación mundial de elementos culturales que conserva aspectos muy importantes con una gran fidelidad. Estos ejemplos muestran que, más allá de la simple replicación física, hay diversos niveles superiores de semejanza abstracta que pueden conservar la información a través de sucesos de selección y, de este modo, permiten que se produzca la evolución cultural por selección natural sin sucumbir a la catástrofe de errores descrita en el capítulo 7. West Side Story es un descendiente de Romeo y Julieta, no porque la secuencia de caracteres impresos o de sonidos que forman la obra de Shakespeare estén reproducidos físicamente en la primera, sino porque se copia el conjunto de situaciones y relaciones representadas; se trata de una replicación de carácter exclusivamente semántico. Parafraseando la familiar advertencia: cualquier parecido entre los

personajes de West Side Story y los personajes de Romeo y Julieta no es pura coincidencia. Si os preguntáis si una replicación semántica así «cuenta» como replicación memética, recordad que los memes son estructuras informacionales habitualmente valiosas (es decir, que vale la pena copiar) y que las leyes de derechos de autor se han concebido y perfeccionado para proteger ese valor. No solo las traducciones, más o menos fieles, cuentan como replicaciones meméticas, sino también los resúmenes, las adaptaciones cinematográficas, las obras y óperas basadas en novelas e incluso los videojuegos. Teniendo en cuenta las leyes de derechos de autor, no puedes partir de una novela sobre unos personajes ficticios en el Boston del siglo XIX, cambiarlo por el París del siglo XVIII y pensar que te puedes salir con la tuya. Es un caso claro de plagio, de robo de un diseño, por mucho que el objeto físico que cuenta como la copia no guarde ninguna semblanza en sus propiedades físicas con el objeto original.

Lo que resulta especialmente importante en este estudio de los memes es que algunos de estos aspectos de alto nivel dependen, en efecto, de la comprensión, no solo de la competencia a la hora de hacer copias, por más que se basen en y dependan de sistemas con una buena competencia de copia que no necesita comprensión. De hecho, desde este punto de vista, podemos ver que la copia de alta fidelidad del ADN, nuestro modelo principal de replicación, destaca como caso extremo de ceguera en la replicación. Para obtener una replicación de alta fidelidad a escala molecular, usando estructuras macromoleculares (polimerasas y ribosomas, por ejemplo) como «lectores» o «copistas», se tiene que confiar en un proceso de copia en extremo simple, casi un «reconocimiento» átomo a átomo y su posterior duplicación. A escalas más altas, con «lectores» más sofisticados y competentes, se pueden crear sistemas que toleren más variación física: un ejemplo destacado son las palabras, pero hay otros. Lso balies ed retlas es peudne reglarr lfácmeinet y l myr d prsns n tndr prblms pr lr st frs. Turing captó la importancia de basar su gran invención en un sistema de reconocimiento tan ciego como fuera posible: elecciones binarias entre 0 y 1; orificio o no orificio en un trozo de papel; voltaje alto o voltaje bajo entre dos puntos de un circuito. En principio, la transmisión de alta fidelidad se puede lograr con un código binario que use cartas de amor para los ceros y amenazas de muerte para los unos, pero entonces se necesitarían lectores con comprensión para obtener una transmisión fiable.[1]

Este sistema es posible en principio, como muchos otros sistemas de digitalización, pero ¿realmente tienen algún papel en la cultura humana? En realidad, sí. Los programadores informáticos usan un excelente término en su jerga: el thinko. Un thinko es como una errata tipográfica (typo, en inglés) pero a un nivel superior, semántico; no un error de escritura, sino un error de pensamiento. Escribir PRITN en lugar de PRINT es una errata; olvidarse de colocar corchetes o asteriscos (o lo que corresponda en cada lenguaje de programación) para delimitar un comentario es un thinko, como también lo es definir un operador binario cuando en realidad necesitas uno ternario. El famoso error del milenio, el Y2K, que no permitía fechas que no empezaran por «19» en las estructuras de datos no era una errata, era un thinko. Un thinko es un error evidente en cualquier esfuerzo en que los objetivos asumidos para la empresa requieran determinadas «prácticas recomendadas» identificables. Los errores (bugs) en los programas informáticos pueden atribuirse a erratas tipográficas en el código fuente, pero acostumbran a ser thinkos (la mayoría de erratas tipográficas son detectadas por el compilador, que las devuelve al programador para que las corrija antes de construir el código ejecutable). Si los expertos no pueden identificar con claridad si algo no cumple con las expectativas, quizá no sea un thinko sino un lapsus lamentable (un thinko es como el concepto de error en el béisbol: al jugador se le dictamina un error cuando ejecuta una maniobra errónea; cuando no está claro cuál hubiera sido una jugada mejor o cuando una jugada es desafortunada porque no resultó tan espectacular como podría esperarse, no se trata de un error). Lo que importa en la replicación de alta fidelidad es que existan cánones de práctica que especifiquen qué thinkos se pueden corregir de forma rutinaria. Truman Capote rechazó un manuscrito de un cuento diciendo «esto no es escribir, es teclear», pero con toda probabilidad no estaba identificando ningún thinko; estaba presentando una norma de excelencia que unas prácticas recomendadas no podían alcanzar con facilidad.

Las rutinas también son memes, perfeccionados durante generaciones gracias a la replicación diferencial y que se pueden reunir formando prácticas más grandes que los expertos pueden «leer» y «escribir». Elaborar flechas y hachas, encender fuegos, cocinar, tejer, hilar, hacer vasijas, puertas, ruedas y barcas y elaborar artes de pesca son maneras de hacer que se pueden ir corrigiendo generación tras generación gracias a la acción conjunta de unos requisitos físicos simples y las tradiciones locales. Cuando estas tradiciones locales evolucionan a «alfabetos»

sencillos de actos, proporcionan un cúmulo de thinkos, erratas de cocina, erratas de tejeduría******** y otros errores fácilmente corregibles para que las tradiciones se puedan transmitir con fiabilidad, y quizá sin lenguaje (tema sobre el que hablaremos en el capítulo 13). Como es habitual, nadie tiene la necesidad de entender que estos «alfabetos» otorgan esta ventaja, y menos que nadie los memes que son sus beneficiarios directos. Una tradición formada por malos hábitos (hábitos parasitarios que reducen la aptitud de los humanos) que consigue ser digitalizada, si el resto de condiciones son iguales, puede sobrevivir a sus competidores en el concurso de la replicación.

El mundo de la danza ofrece unas interesantes variaciones sobre este mismo tema si comparamos las danzas tradicionales con las coreografías sofisticadas. Las danzas tradicionales tienen un «alfabeto» bien conocido por los bailarines. En las contradanzas y las square dances, por ejemplo, hay una serie de pasos y movimientos básicos con nombres propios que todo el mundo conoce (honor your partner, honor your corner, etc.); el director de baile puede crear una nueva danza al momento sin más que cambiar el orden de esos pasos o sus valores.******** Es un lenguaje de programación muy simple para la construcción de una contradanza. Como resultado de este sistema digitalizado, un grupo de personas que no haya bailado una virginia reel durante décadas y que no recuerde bien los pasos, puede recrear la danza con unos pocos ensayos (replicaciones), ya que la ley de la mayoría se llevará por delante las erratas y los thinkos (pasos equivocados) y recuperará la «ortografía» original. No habrá dos parejas que hagan exactamente igual el paso de alemanda a la izquierda, pero sus variantes serán aceptables para el baile completo. Esta «votación» se ha reinventado una y otra vez en la cultura humana y es una forma fiable de acentuar la fidelidad de la transmisión por parte de memorias individuales poco fiables y de baja fidelidad.

El canto monofónico es omnipresente en las religiones tradicionales y otras ceremonias y, de forma parecida, permite reparar la memoria de los cantantes, ninguno de los cuales puede lograr sin acompañamiento una copia fiel de la ejecución del año anterior. Cuando en el siglo XVIII se inventaron los cronómetros marinos, los navegantes aprendieron a llevar tres en sus largos viajes, y no solo dos, para así poder usar el truco de la ley de la mayoría (para entonces un meme domesticado y aceptado). El poder replicador del canto tribal no tiene que ser reconocido ni apreciado por los cantantes, aunque algunos miembros pensativos pueden haberlo

captado. Las motivaciones flotantes han sido tan ubicuas en la evolución cultural como en la genética, y pueden ser igualmente informativas para nosotros en cuanto consigamos sofocar nuestra costumbre de atribuir comprensión a cualquier organismo, o meme, que haga alguna cosa inteligente.

A diferencia de las danzas tradicionales, las coreografías sofisticadas exigen una sutileza que obliga a un sistema de recetas más expresivo y, aunque el cine y el vídeo han «solucionado» este problema mediante fuerza bruta (de forma parecida a cómo la música grabada puede «solucionar» el problema de una partitura desaparecida), la coreografía, a pesar de su nombre, no ha logrado encontrar hasta la fecha un sistema para representar sus movimientos y pasos lo bastante eficaz como para quedar bien establecido entre bailarines y coreógrafos. La notación de Laban, inventada en la década de 1920 por Rudolf Laban, tiene un leal grupo de aficionados, pero no ha logrado imponerse como lingua franca en el mundo de la danza.

Con lo que hemos dicho, resulta tentador ver una transición gradual en estos tres pasos:

Sincronización rítmica «infecciosa» entre miembros de una tribu que bailan, hacen ruidos y vocalizan en encuentros improvisados, repitiendo sus movimientos preferidos e imitando los de los demás; he aquí el origen sinantrópico de la danza, sin directores, sin comentaristas, sin coreógrafos.

Rituales más conscientes, que necesitan ensayos y un proceso de instrucción y corrección intencionado; estamos ante la domesticación de la danza gracias a un control cuidadoso de la reproducción.

Coreografías profesionales actuales, con los coreógrafos como ingenieros meméticos que diseñan inteligentemente sus obras de arte y las presentan ante el mundo cultural con la esperanza de que se reproduzcan.

Según este punto de vista, las maneras originales de bailar eran memes que no pertenecían a nadie, que evolucionaban a ciegas aprovechándose de las particularidades de la locomoción, la percepción y la excitación emocional humanas. ¿Para qué servían? Para prosperar en compañía humana, hábitos que se propagaban porque podían propagarse, como el resfriado común.

Sin duda, en este proceso sería ventajoso que los memes originales de la danza ofrecieran algún tipo de beneficio a los genes de los protobailarines, pero este mutualismo podría haber evolucionado gradualmente a partir de una relación de comensalismo o incluso de parasitismo, mientras los memes en competición luchaban por tener algo de tiempo disponible en los cuerpos aptos. Las malas costumbres infecciosas pueden ser difíciles de erradicar, pero si se convierten en hábitos útiles, aumentarán sus perspectivas de reproducción; una vez identificados, vagamente al principio (la «selección inconsciente» de Darwin) y luego ya conscientemente (la «selección metódica» de Darwin), sus anfitriones garantizarán su reproducción, de modo que los memes se podrán relajar, volverse menos apasionantes, menos irresistibles, menos cautivadores, menos vívidos y menos inolvidables, precisamente porque se han vuelto muy útiles. Los cerebros de los animales domesticados siempre son más pequeños que los de sus parientes salvajes más cercanos (o lo tomas o lo dejas) y los animales domesticados tienen una vida comparativamente plácida, protegidos de los depredadores y del hambre y con parejas y apareamiento asegurados. El resultado de todo esto es que para que algo soso se propague, sus anfitriones tendrán que considerarlo especialmente útil o valioso y, por tanto, tener interés en que se reproduzca, inculcándolo a través de la formación. Algunos ejemplos que se me ocurren son la contabilidad por partida doble o la trigonometría. Ser considerado valioso por parte de anfitriones influyentes es una adaptación habitual de los memes, y entre los ejemplos más radicales podemos citar la soporífera música «seria» atonal y algunas versiones del arte conceptual contemporáneo que, cual gallinas ponedoras, están tan domesticadas que se extinguirían en una generación si no fuera por la ayuda de sus adinerados patrocinadores.

Así pues, tal como Richerson y Boyd, Sperber y otros han señalado, aunque es cierto que muchos aspectos importantes de la cultura humana muestran cambios a lo largo del tiempo que no son producto directo exclusivo de sistemas de transmisión de alta fidelidad de entidades

informacionales discretas o pseudogenéticas, estos sistemas (no solo palabras, también la música, la danza, la artesanía y otras tradiciones de ofrecimientos) proporcionan diversas rutas por las que se puede transportar la información cultural de una generación a la siguientes con la fidelidad suficiente para que mutaciones graduales relativamente ciegas provoquen una acumulación de mejoras sin necesitar más que una mínima comprensión. Tal como vimos en el caso de las barcas, los propios artefactos proporcionan normas de corrección. Como regla general aproximada, cualquier artefacto que se encuentre en abundancia y que muestre señales de uso es un buen lo-que-sea. Siguiendo esta regla, a menudo se pueden distinguir los buenos artefactos de los mediocres sin saber exactamente por qué los buenos son buenos; y, a partir de ello, ¡copia los buenos, por supuesto! La brillante idea de Darwin de selección inconsciente como una transición gradual hacia la domesticación también resulta importante en la evolución cultural. Nuestros ancestros ignoraban «automáticamente» a los patitos feos y las manzanas podridas, y en cada caso, el resultado fue una mejora gradual de la descendencia (en relación con los gustos y necesidades humanos).

LOS MEMES, A DIFERENCIA DE LOS GENES, NO TIENEN ALELOS CON LOS QUE COMPETIR EN UN «LOCUS»

En el ADN, los genes se representan como largas secuencias de nucleótidos (A, C, G y T), que permiten comparar diferentes genomas e identificar los lugares (locus) en que son iguales y los lugares en donde difieren. En la representación física de los memes no existe nada parecido (aún); no hay ningún código básico común para representar un meme concreto ni sus competidores cuando se producen sus casos. Por ejemplo, Benjamin Franklin inventó las lentes bifocales y, sin duda, en su cerebro había alguna estructura neuronal que encarnaba o representaba su creación, el concepto de lentes bifocales. Cuando publicitó su invención, otros recogieron la idea, pero no tenemos ninguna buena razón para suponer que las estructuras neuronales de los cerebros de estos otros representaron el concepto de lentes bifocales del mismo modo que Franklin. Para algunos críticos esto equivale a admitir que no

existe ningún meme lentes bifocales, que los memes no existen y que la teoría de los memes es una mala idea. De momento, lo que significa es que no podemos usar un código común para identificar locus en diferentes cerebros y ver si en el locus de lentes bifocales hay la misma representación del meme o una diferente, del mismo modo que una prueba genética puede determinar si una persona tiene el alelo (variante) de la corea de Huntington o la enfermedad de Tay-Sachs.

Ahora bien, ya hemos visto que hay sistemas «alfabéticos» que se pueden usar para las palabras, la música y muchas otras familias de memes, que crean buzones, casilleros o locus para una clase algo diferente de competidores alélicos. Para algunas palabras, hay pronunciaciones competidoras (cóctel o coctel; mísil o misil), mientras que con los préstamos léxicos procedentes de otros idiomas, suele haber una competición entre una pronunciación vagamente parecida al original y una pronunciación más naturalizada al idioma de destino. Así, los ingleses home run y wifi a menudo se transforman en jonrón y güifi en español, y ¿qué pensará el camarero de una trattoria romana cuando le pidas una «picsa ai fungi porzini» y unas «tagliatelas a la bolocnesa»? También podemos mantener la pronunciación fija y mirar significados competidores en un locus fonológico. Huésped es un autoantónimo, que significa una cosa y su opuesto, aunque en este caso el significado que equivale a anfitrión ya casi se ha extinguido (excepto en biología); lo mismo sucede con enervar, donde el significado de ‘debilitar, apaciguar’ es cada vez menos usado. Vaina ha acabado teniendo significados habituales muy diferentes a ambos lados del Atlántico, mientras que es muy posible que hoy en día pocas personas conozcan el significado original de badulaque, término que, gracias al doblaje de una conocida serie de animación, en España ha pasado a significar una cosa totalmente distinta.

Como sucede en los genes, las mutaciones son errores de transmisión, pero a veces un error es una mejora afortunada; lo que empezó como un error, se ha convertido en un rasgo. También hay casos como el del término inglés skyline, que según el Oxford English Dictionary, originalmente significaba «en una escena perceptible visualmente, la línea que delimita el cielo de los elementos terrestres», pero que, poco a poco, ha restringido su significado solo a las situaciones urbanas (Richard, próxima publicación). Una cadena montañosa en lontananza, recortada contra el azul del cielo, en inglés ya no se puede denominar

con propiedad un skyline.[2] Otra posibilidad es fijar el significado y echar una mirada a los locus semánticos en los que se libra la batalla de los sinónimos, con núcleos de resistencia regionales y términos nostálgicos anticuados camino de la obsolescencia (en la década de los sesenta del siglo pasado, las fiestas eran guateques, y seguro que los jóvenes de hoy en día ya no salen de marcha ni a ligar como hacían sus padres). Y no es que estos cambios de significado y de pronunciación (y también gramaticales) pasaran desapercibidos hasta la aparición de los memes; de hecho, los lingüistas históricos los han estudiado con rigurosidad desde hace siglos, pero estos investigadores a menudo se han visto entorpecidos por suposiciones acerca de significados esenciales, creando controversias artificiales que tienen un sorprendente parecido con las batallas entre taxonomistas acerca de especies, géneros, variedades y subespecies, solo apaciguadas tras Darwin[3] (sí, increíble realmente significa 'que no se puede creer', y sí, las aves realmente son dinosaurios y los perros son lobos).

En música, podemos considerar que cada canción popular es un locus, con las diversas versiones de diferentes grupos peleándose por su dominio. Desde este punto de vista, vemos que, a menudo, la versión de éxito de una canción, que se convierte casi en canónica y se fija, es una mutación, un alelo rival de la versión original. Algunas versiones originales de cantautores famosos han quedado suplantadas por versiones posteriores (¿alguien ha oído la versión original de Kris Kristofferson de «Me and Bobby McGee»?); estos cantautores cobran royalties por las versiones de los otros cantantes, aunque el valor añadido de posteriores intérpretes les proporciona la mayor parte de las ventas. La existencia de tecnología de grabación de sonido ofrece la posibilidad de disponer de versiones fijas y canónicas de una canción, lo que solidifica la identidad física de cada versión. Experimentos llevados a cabo por Daniel Levitin (1994, 1996) aprovecharon esta regularidad para probar la memoria de diversos individuos por lo que respecta al tempo y tono de diversas canciones populares. Mientras que Greensleeves no tiene ninguna tonalidad ni tempo canónicos, «Hey, Jude» y «Satisfaction» sí, y sus admiradores eran muy buenos al tararear o cantar estas canciones de memoria con un tempo y tono muy cercanos al correcto.

Cuando pensamos si los memes se replican con fidelidad suficiente para permitir la acumulación de adaptaciones, debemos incluir las tecnologías de mejora en la cultura, tal como hacemos con los genes. Buena parte de

la I+D evolutiva se invirtió en mejorar la maquinaria de replicación del ADN durante los primeros mil millones de años de la vida, más o menos. La invención de la escritura ha impulsado de forma parecida la fidelidad de la transmisión lingüística, fruto de muchas mentes en muchos lugares durante varios milenios. Bien pocos «inventores» de la escritura tuvieron (o necesitaron tener) una visión clara de las especificaciones de la máquina que estaban creando ni del problema que estaban solucionando con tanta brillantez (comparemos su trabajo con la invención de la computadora por parte de Turing o el diseño de arriba abajo del programa controlador del ascensor). Tal como vimos en el capítulo 6, las leyes de derechos de autor requieren que haya un registro más o menos permanente (como un texto escrito o una grabación de audio), por claros motivos de pruebas legales, pero la capacidad de traducir o transducir una invención en un medio así tiene un impacto más profundo que la mera protección legal. «El medio escrito permite una mayor complejidad porque las palabras en una página no se las lleva el viento como las del habla, y pueden examinarse una y otra vez hasta entender lo que el autor pretendía decir» (Hurford 2014, p. 149). Esto nos otorga la capacidad de descargar en el mundo una frase, un verso, una saga o una proclama y así quitarnos el peso de encima de tener que confiar en nuestra limitada memoria y darnos el tiempo de dar vueltas y más vueltas a un fragmento de texto sin tener que preocuparnos de si se modificará poco a poco en nuestros ensayos o de si se diluirá de nuestra memoria[4] (con mucha práctica, puedes hacer multiplicaciones de tres dígitos mentalmente, pero sin duda es mucho más fácil con lápiz y papel).

Gracias a nuevas innovaciones, es posible que actualmente nos hallemos a las puertas de una transición importante en la evolución cultural. No hay ningún ADN para la cultura, hoy en día, pero el HTML (el lenguaje en que se basa la representación de contenidos en Internet) o sus descendientes pueden acabar siendo tan dominantes en el futuro que pocos memes podrán competir por «ojos» y «oídos» en este atestado mundo de información sin estar representados en HTML. Ahora hay bots y aplicaciones que pueden buscar contenidos e identificar y distinguir contenido (como Shazam, la aplicación para teléfonos inteligentes); si sus descendientes empiezan a hacer lo que podríamos considerar juicios de valor (o transmisión sesgada, en palabras de Richerson y Boyd 2004), alentando la replicación diferencial de memes sin ninguna dependencia de los cerebros, ojos u oídos humanos, los memes del futuro podrán medrar sin intervención humana directa; todavía serán sinantrópicos,

como la golondrina común y el vencejo de chimenea, pero dependerán de los servicios del nicho tecnológico construido en el siglo XXI por los seres humanos. Susan Blackmore (2010) había bautizado a estos memes copiados digitalmente como temes, pero hace poco ha decidido denominarlos tremes (correspondencia personal, 2016) (en el último capítulo volveremos sobre el tema).

No todos los procesos de cambio cultural muestran rasgos análogos a la competencia de alelos en los locus, pero en cualquier caso, este rasgo solo es una de las dimensiones en las que pueden variar los procesos evolutivos, algunas más darwinianas que otras. Darwin no disponía del concepto de locus con alelos y tampoco lo necesitó para establecer los aspectos básicos de su teoría de la evolución por selección natural (en el capítulo 13 echaremos un vistazo a otras dimensiones de variación, dispuestas en uno de los espacios darwinianos de Godfrey-Smith).

LOS MEMES NO AÑADEN NADA A LO QUE YA SABEMOS SOBRE LA CULTURA

Esta afirmación nos dice que los potenciales memetistas se han limitado a acaparar todas las categorías, relaciones y fenómenos descubiertos, descritos y explicados por los teóricos culturales de escuelas de pensamiento más tradicionales y abiológicas y a renombrar todo este campo como memética. Según esta objeción, no añadimos nada relevante al pasar de hablar acerca de ideas, creencias y tradiciones a hablar acerca de memes (con su pátina pseudobiológica y pseudocientífica). ¿Dónde están las nuevas percepciones, correcciones, refinamientos que podrían motivar o justificar este golpe de estado? Los memetistas no hacen más que reinventar la rueda (espero haber hecho justicia a la inflexible indignación con la que se suele expresar esta objeción).

Hay parte de verdad en esta acusación y es cierto que a veces los memetistas (entre los que me incluyo) hemos presentado como novedosos resultados que, en realidad, los teóricos culturales tradicionales ya habían formulado y asimilado hace mucho tiempo. Para

un teórico de la cultura (un historiador, un antropólogo o un sociólogo, por ejemplo), debe ser especialmente irritante aguantar una explicación memética elemental de algún fenómeno, explicación que retoma inconscientemente explicaciones mucho más detalladas y respaldadas que él mismo había elaborado años atrás. Ahora bien, sería sorprendente descubrir que los teóricos culturales anteriores habían planteado teorías incorrectas o que no habían sido capaces de distinguir las diferencias identificadas por los memetistas, de modo que para un memetista no debería ser motivo de vergüenza admitir que, en algunos aspectos, los teóricos culturales se las habían arreglado bastante bien con sus proyectos explicativos sin memética.

Los memetistas deberían buscar y atesorar los resultados recogidos por los investigadores tradicionales de la cultura, siguiendo el ejemplo de Darwin, que recopiló los abundantes datos de historia natural que descubrió gracias a la extensa correspondencia que mantuvo con sus colegas. Los datos recogidos y ordenados acerca de incontables aspectos de plantas y animales por parte de investigadores predarwinianos tenían la rara virtud de ser teóricamente impolutos, al no haber sido recogidos por darwinianos apasionados (o por antidarwinianos apasionados) cuyos prejuicios podrían haber nublado sus opiniones. Igualmente, los estudios y teorías culturales prememéticos no darwinianos deben valorarse por la misma razón: la recopilación de datos (que en las ciencias sociales siempre está en riesgo de quedar manchada por los sesgos del observador) no se ha efectuado con la intención de respaldar la memética. La memética sería una dudosa candidata a buena teoría de la evolución cultural si no pudiera englobar una buena parte de las teorías anteriores adaptando pequeños aspectos terminológicos para reformularlos desde el punto de vista de los memes.

Pero, además, la memética debe poder ofrecer avances genuinos en perspectiva. Ya he esbozado lo que creo que son sus contribuciones principales, junto con los puntos débiles en las teorías tradicionales de la cultura. Las culturas rebosan de elementos bien diseñados y los teóricos culturales o bien han sobredotado a las personas con comprensión, habilidades inventivas y genialidad o bien (apreciando que las personas, como las mariposas o los antílopes, pueden ser los beneficiarios de disposiciones que no necesitan comprender) han renunciado a la responsabilidad de explicar cómo puede haber surgido la I+D necesaria. La evolución genética («instintos») no puede actuar lo bastante rápido

para poder encargarse de este proceso, lo que nos deja con un enorme abismo que la memética debe poder rellenar, y más teniendo en cuenta que ninguna idea procedente de los enfoques tradicionales sobre la cultura lo puede hacer.

El punto de vista memético también es valioso para «depsicologizar» la propagación de innovaciones (buenas y malas). Los enfoques tradicionales en evolución cultural (en el sentido neutro de cambio cultural a lo largo del tiempo), como la «historia de las ideas» y la antropología cultural, consideran que las personas son, ante todo, seres que perciben, creen, recuerdan, planean, saben, comprenden y aperciben. Una persona en coma o durmiendo no es transmisor/transductor de cultura en ninguna situación, de modo que la conjetura natural es que las innovaciones culturales se aperciben y luego se adoptan (a menudo); así, la idea es que las innovaciones se adoptan porque se han valorado, se han considerado valiosas o simplemente se han deseado (también se pueden adoptar por error, sin duda). Todo esto encaja a la perfección (demasiado) con la visión predeterminada de las personas como agentes racionales, sistemas intencionales cuyo comportamiento puede predecirse en principio atribuyéndoles creencias, deseos y racionalidad. A su vez, esto lleva a una versión u otra del modelo económico de conservación y transmisión cultural: los «productos» culturales que se consideran valiosos se conservan y mantienen y o bien se legan a la siguiente generación o bien se venden al mejor postor. Pero buena parte de la innovación cultural se produce mediante lo que podríamos llamar ajustes subliminales en largos periodos, sin necesidad de que nadie se dé cuenta ni de que nadie los apruebe. Como los virus, los memes se pueden propagar de anfitrión a anfitrión sin que estos se den cuenta. Los efectos a escala poblacional se pueden acumular, ignorados por los miembros de esa población. Estos cambios acumulados a menudo se identifican retrospectivamente, como cuando a una comunidad de emigrantes llega una nueva persona procedente del país de origen, con un modo de hablar a la vez familiar y extraño («¡Cierto! Recuerdo que solíamos hablar así»). Estos cambios subliminales no solo se dan en la pronunciación y en el significado de las palabras; en principio, las actitudes, valores morales y las particularidades más emblemáticas de una cultura pueden suavizarse, endurecerse, erosionarse o quebrarse a un ritmo demasiado lento para percibirlo. La evolución cultural es rapidísima comparada con la evolución genética pero aun así todavía puede ser demasiado gradual para que una observación casual la identifique.

También hay las «variantes culturales patológicas». Muchos fenómenos culturales persistentes son maladaptativos, y lo cierto es que ninguna teoría que equipare las innovaciones culturales con las genéticas («adaptaciones transmitidas por otros medios») es capaz de explicarlos. En el capítulo «Culture is Maladaptive» de Richerson y Boyd (2005) no se presenta una defensa de la memética, pero sí que adopta el punto de vista de los memes para explicar una serie de fenómenos aparentemente desconcertantes, como el auge (hasta la fecha) de grupos anabaptistas como los amish. A pesar de sus recelos sobre algunos aspectos de los memes (véase, p. ej. p. 6), aceptan que «el efecto del meme egoísta es bastante robusto» (p. 154).

Hasta el momento, solo los anabaptistas y unos pocos grupos similares, como los judíos ultraortodoxos, parecen tener una buena resistencia a las infecciones de la modernidad. El anabaptismo es como un robusto cayac que navega en el turbulento mar de la modernidad; parece frágil pero sobrevive porque no tiene ninguna vía de agua, a pesar de los enormes esfuerzos a que está sometido. Una vía de agua grave por algún sitio y se acabó. Es imposible predecir el futuro o futuros evolutivos del anabaptismo. Mientras tanto, ¡solo nos queda admirar la belleza de su diseño!

LA PSEUDOCIENCIA DE LA MEMÉTICA NO ES PREDICTIVA

De algún modo, esta afirmación es cierta, pero no es una objeción a la memética. La teoría de la evolución genética tampoco es predictiva, en el sentido de que nos permita describir el futuro de Homo sapiens (o del bacalao, del oso polar o de los abetos) con un mínimo grado de certeza. Y hay una razón profunda para que sea así: los procesos evolutivos darwinianos son amplificadores de ruido. Tal como señalé en el capítulo 1, un rasgo curioso de la evolución por selección natural es que depende de forma crucial de sucesos que no acontecen casi nunca. Así, la teoría evolutiva no puede predecir esos sucesos que se dan una sola vez de mil

millones y que se irán amplificando para dar lugar a nuevas especies, nuevos genes, nuevas adaptaciones y, en consecuencia, no puede predecir el futuro excepto de un modo muy condicional: si sucede tal y cual, entonces (a no ser que se amplifique otro suceso impredecible hasta machacar al primero) pasará esto y lo otro. Pero esta especie de predicción al instante no es un requisito necesario para la buena ciencia; la predicción de qué es lo que encontraremos y lo que no encontraremos si excavamos para buscar fósiles (por poner un ejemplo claro) sigue siendo una predicción, aunque hace referencia a sucesos que ocurrieron hace millones de años. La biología evolutiva predice que nunca encontrarás un pájaro con pelo en lugar de plumas, por muy aislada que esté su población en una isla lejana, y que, en una especie de insecto recién descubierta, hallarás secuencias muy concretas de ADN en lugares muy determinados de su genoma.

Así pues, que los memetistas no puedan decir con seguridad qué canciones estarán en el top 20 en 2018 o si se pondrán de moda los dobladillos de los pantalones no es una objeción grave a la ciencia de la memética. Una pregunta más pertinente es si la memética puede proporcionar explicaciones unificadas de los patrones observados por los investigadores de la cultura pero que no se han unido bajo una descripción común o ni tan solo se han podido explicar. Sostengo que el punto de vista de los memes rellena esta gran y problemática laguna entre los instintos transmitidos genéticamente y las invenciones con comprensión, entre animales competentes y diseñadores inte-

ligentes, y la rellena con el único marco teórico que puede explicar sin

milagros la acumulación de buenos diseños: la replicación diferencial de la descendencia. Está por ver si alguna otra teoría evolutiva del cambio cultural puede también rellenar esta laguna.

LOS MEMES NO PUEDEN EXPLICAR

LOS RASGOS CULTURALES; LAS CIENCIAS SOCIALES TRADICIONALES, SÍ

Plantear esta objeción es no entender qué son los memes, y podemos verlo sugiriendo la misma objeción para los genes: «los genes no pueden explicar las adaptaciones (estructuras, órganos, instintos, etc.)». Es cierto; es por ello por lo que necesitamos la biología molecular, la fisiología, la embriología, la etología, la biogeografía insular y todas las demás ramas especializadas de la biología si queremos explicar cómo funcionan adaptaciones concretas y por qué son adaptaciones. También necesitamos otras disciplinas para explicar cómo explotan los parásitos a sus anfitriones, por qué las telarañas son unas trampas eficientes, cómo los castores construyen sus presas, por qué vocalizan las ballenas y tantos otros fenómenos. Del mismo modo, necesitamos la psicología, la antropología, la economía, las ciencias políticas, la historia, la filosofía y la teoría literaria para explicar cómo y por qué los rasgos culturales (buenos y malos) funcionan del modo en que funcionan.

Theodosius Dobzhansky estaba en lo cierto cuando con cierta concisión dijo que «nada en la biología tiene sentido excepto a la luz de la evolución», pero no dijo que la luz de la evolución dé sentido a todo lo que hay en biología. No hay nada en la memética que pueda sustituir el arduo trabajo de recolección de datos y comprobación de patrones de la «ecología conductual» de las personas en la música y las artes, las religiones, las empresas, los ejércitos, las asociaciones, las familias, los equipos, las escuelas, las cocinas y todos los demás contextos de la vida humana. Lo que la memética promete es un marco para que todo esto cobre sentido en algunos de sus aspectos. Nadie nace siendo sacerdote, fontanero o prostituta; y solo con sus genes o con los memes que los infectan no podremos explicar cómo llegaron a ser lo que son. En su punto álgido, la historia natural predarwiniana era una ciencia sistemática y desarrollada, con hipótesis que debían comprobarse y abundantes explicaciones, pero Darwin la enriqueció con un conjunto de preguntas que permitieron ver todas sus generalizaciones bajo una nueva luz. Mi afirmación global es que la perspectiva evolutiva, en general, y la memética por lo que respecta a la cultura, transforma muchos de los enigmas de la vida aparentemente eternos, como el sentido y la conciencia, gracias a vías que resultaban inaccesibles a los que nunca miraban más allá de la imagen manifiesta con la que crecieron y de las disciplinas en las que se formaron.

LA EVOLUCIÓN CULTURAL ES LAMARCKIANA

Esta es una de las «explicaciones» más populares de por qué los enfoques darwinianos sobre la evolución cultural son erróneos, y es una explicación que lleva por bandera la confusión (y la desesperación). Es muy posible que los que plantean esta «acusación» tengan en mente la doctrina propuesta por Jean-Baptiste Lamarck, predecesor de Darwin y según el cual las características adquiridas por un individuo durante su vida se pueden transmitir genéticamente a su descendencia: los potentes brazos del herrero, adquiridos gracias a su extenuante uso, pasan a su descendencia a través de los genes (y no mediante la imposición de unos rigurosos hábitos de ejercicio a sus hijos). Igualmente, según este punto de vista lamarckiano, el miedo a los humanos inculcado en un perro a causa de un encuentro con un dueño cruel puede pasar a ser un miedo instintivo en las crías del perro, las cuales no necesitan haber experimentado la crueldad para preverla o temerla. Pero las cosas no van así. No hay ningún modo de que un carácter adquirido pueda ajustar los genes de un organismo y que, así, el carácter pase genéticamente a la siguiente generación (para ello el organismo tendría que efectuar un proceso de ingeniería inversa del cambio somático [corporal], descubrir la receta genética para que esa novedad se diera durante el desarrollo y, finalmente, ajustar la secuencia de ADN correspondiente en los óvulos o el esperma). Esta idea en verdad herética no solo fue adoptada por Lamarck, pues diversas variantes estuvieron en boga durante el siglo XIX. También Darwin coqueteó con una de sus versiones, pero actualmente queda totalmente refutada por la teoría evolutiva neodarwiniana. No cabe duda de que hay diversos fenómenos reales que parecen bastante lamarckianos, porque implican una cierta transmisión de caracteres adquiridos hacia la descendencia, pero no genéticamente. Por ejemplo, un perro temeroso podría transmitir el miedo a sus crías a través de una feromona («el olor del miedo») o mediante una hormona compartida a través de la placenta, pero no a través de sus genes. El efecto Baldwin también parece lamarckiano en algunos aspectos, ya que los comportamientos adquiridos por una generación pueden crear una presión selectiva que favorezca a la descendencia con una mayor tendencia o talento para adquirir esos comportamientos, lo que, a la larga, los convertirá en «instintivos». Y, por si fuera poco, a esto se añaden los temas candentes actuales: la complejidad del desarrollo

descubierta por los biólogos evolutivos del desarrollo y, muy en particular, los caracteres epigenéticos. En 2008 se logró una definición de consenso de carácter epigenético: «un fenotipo heredable de modo estable producto de cambios en un cromosoma sin alteración de la secuencia de ADN» (Berger et al. 2009). Se trata de un tema candente por varias razones: los investigadores han descubierto unos giros inesperados y fascinantes a escala molecular en la inacabada saga de la teoría evolutiva, algunos de los cuales han sido aclamados como revolucionarios, pues aparentemente ponen en tela de juicio la ortodoxia neodarwiniana; en consecuencia, algunas personas temerosas de Darwin, fuera del ámbito de la biología, se han despabilado, animados por lo que han oído y han propagado el falso rumor de que el descubrimiento de la herencia lamarckiana ha «refutado» la evolución por selección natural. Tonterías.

Presuntamente, algunos críticos de los memes actúan con la impresión de que si la evolución cultural es lamarckiana, no puede ser darwiniana (¡uf!). Los críticos están en lo cierto que la transmisión cultural permite que los caracteres adquiridos por los progenitores se inculquen a los hijos (mediante ejemplos, instrucción o castigos), pero esto no demuestra que la evolución cultural no se logre mediante selección natural darwiniana; ni por asomo. Al fin y al cabo, los padres también pueden pasar a sus hijos los gérmenes, microbios y virus adquiridos y todos estos fenómenos están sometidos, sin atisbo de duda, a la selección natural darwiniana (no lamarckiana). Quizá estos críticos olvidan que en la evolución memética es la aptitud de los memes lo que está en juego, no la aptitud de sus anfitriones. La pregunta del lamarckismo debería ser si los caracteres adquiridos por los memes se pueden transmitir a su descendencia; y aquí, como no existe la distinción entre genotipo y fenotipo que imposibilita la transmisión lamarckiana, esta ya no sería una herejía, sino solo una variedad alternativa de selección natural. Después de todo, los memes no tienen genes (véase, p. ej., la figura 7.2 en la p. 135, el espacio darwiniano con la dimensión G, que representa la distinción germinal/somática: las diferentes posiciones en el eje G admiten la herencia lamarckiana, entendida en ese sentido). En un virus, por ejemplo, no hay una distinción clara entre mutaciones en la línea germinal y caracteres adquiridos; el replicador y el interactuador son una misma cosa (para la mayoría de propósitos).

Hay otro aspecto, más interesante, en que la evolución cultural se podría considerar lamarckiana, aspecto que se puede ver más fácilmente si consideramos nuestros memes paradigmáticos: las palabras. Tal como vimos en el capítulo 9, antes de que un bebé empiece a intentar imitar una palabra, tiene que haber oído unos seis casos (seis «progenitores», no los habituales dos —o uno en el caso de reproducción asexual—). Pero ahora supongamos que yo, como adulto, aprendo una palabra nueva al oírla decir a alguien en un contexto en que queda claro su significado; así, guardo una copia de ese caso como un nuevo elemento léxico de mi cerebro. La estructura informacional de mi cabeza es un descendiente de ese caso, y cada vez que la repito, estos nuevos casos son otros tantos descendientes (nietos y biznietos, diríamos) del caso que había expresado esa persona. Pero supongamos que antes de que yo exprese la palabra en voz alta vuelvo a oírla decir a esa persona, o la oigo de los labios de otros hablantes; ahora tendré varios modelos de mi nueva palabra, varios padres, por así decirlo.

Podríamos considerar que mi nueva palabra, como la del bebé, tiene varios padres que han contribuido a fijar la estructura informacional que es su descendencia; o podríamos considerar que la primera expresión oída es el único progenitor (como los virus, que se reproducen asexualmente) y que todos los otros casos que han ayudado a fijar los rasgos estables de mi versión de la palabra no son progenitores, sino solo influidores. Al fin y al cabo, tras haber oído el primer caso, mi memoria y mi sistema perceptivo ya está preparado para reconocer casos adicionales y cuando oigo casos posteriores emitidos por otros, pueden inducirme a hacer nuevos casos propios, pero no se tratará de una imitación, sino de una «reproducción desencadenada» (Sperber 2000). Pero desencadenar la reproducción, si no es exactamente lo mismo que ser padre, sí que es más o menos hacer de comadrona, contribuyendo (quizá con un cierto sesgo) a la posterior reproducción del tipo. Mi estructura informacional de la palabra se podrá ajustar al encontrarme con estos casos posteriores de la palabra, que, a su vez, pueden provocar cambios en los rasgos a heredar por cualquier descendencia posterior de mi versión. Esto sería una especie de lamarckismo a la escala de los memes, una de las muchas variantes de la selección natural en poblaciones darwinianas. Alternativamente podríamos considerar que las palabras, y los memes en general, son el resultado de procesos de reproducción variables y prolongados en el tiempo (como si la contribución del padre no se produjera en el momento de la fecundación,

sino más tarde, después de que la madre haya dado a luz), una variación imaginable de nuestro modo normal de reproducción sexual.

En definitiva, hay varios modos en que podemos considerar que la evolución cultural es lamarckiana sin que por ello tengamos que erigir una muralla contra las incursiones imperialistas del temido darwinismo hacia los sagrados templos de la cultura. Steven Pinker (no precisamente un amigo de los memes) ha admitido con candor que «decir que la evolución cultural es lamarckiana es confesar que no se tiene ni idea de cómo funciona». Sin embargo, prosigue para decir unas cuantas cosas sobre cómo cree que funciona:

Los rasgos más destacados de los productos culturales, a saber, su ingenio, belleza y verdad (análogos al complejo diseño adaptativo de los organismos) provienen de los cálculos mentales que «dirigen» —es decir, inventan— las «mutaciones» y que «adquieren» —es decir, entienden y comprenden— los «caracteres». (1997)

Esto expresa perfectamente el punto de vista tradicional de que lo que explica las mejoras (el «ingenio, belleza y verdad») que observamos en los elementos culturales es la comprensión por parte de inventores, por parte de diseñadores inteligentes. En efecto, algunas de las maravillas de la cultura se pueden atribuir a la genialidad de sus inventores, pero muchas menos de lo que se suele creer, y todo se basa en los cimientos de un buen diseño formado durante milenios por un ejército de memes sin comprensión que compiten unos con otros por el tiempo de repetición disponible en los cerebros.

A partir de esta reflexión de las objeciones a los memes surgen varios temas. En primer lugar, el problema no debería basarse en si Dawkins, al presentar el concepto, consiguió articular la mejor definición, o si yo o cualquier otra persona ha tenido éxito en elaborar una formulación a prueba de balas. El problema radica (como suele pasar en ciencia) en si hay conceptos y perspectivas valiosas que surgen de las investigaciones realizadas hasta la fecha. La analogía con los genes es fecunda pero parcial, y acaso la principal ventaja del punto de vista de los memes es que propone preguntas sobre fenómenos culturales que quizá no nos

hubiéramos planteado nunca, como «¿es x el resultado de diseño inteligente? ¿Es x un bien que vale la pena conservar y legar o solo un pedazo de basura parasitaria? ¿Hay alternativas (alelos) a x que han sido derrotadas?».

En segundo lugar, los memes compiten con las descripciones tradicionales del cambio cultural solo cuando estas descripciones atribuyen comprensión a las personas (o misteriosas fuerzas sociales) pero para las cuales, en realidad, no hay pruebas de nada más que la excelencia de los diseños. Así, tal como veremos en el capítulo 13, la perspectiva memética ofrece un terreno de juego equilibrado, donde se pueden encontrar todos los grados y tipos de comprensión humana.

Aún no existe una ciencia formalizada de la memética, y quizá nunca la habrá, aunque se han efectuado unos cuantos intentos pioneros. Pero, del mismo modo, las partes de la teoría evolutiva en biología que han fructificado con un tratamiento matemático son solo eso, partes, y Darwin se las apañó sin ellas. Darwin también aplazó, sabiamente, las indagaciones sobre cómo se inició la evolución hasta haber descrito con detalle su funcionamiento a medio viaje: «descendencia con modificación». Siguiendo su ejemplo, ahora que hemos visto cómo las palabras y otros memes pueden descender, con modificaciones, de sus antecesores, podemos retomar la delicada pregunta de cómo se originó en nuestra especie la evolución cultural, en general, y el lenguaje, en particular.

******** En el original, cookos (de to cook, ‘cocinar’) y weavos (de to weave, ‘tejer’), que mantiene el juego de palabras con typo y thinko. (N. del t.)

******** La square dance es un tipo de baile tradicional para cuatro parejas formando un cuadrado, muy popular en Estados Unidos que, como dice el autor, «todo el mundo conoce», y cuyo origen se halla en las contradanzas europeas. El ejemplo planteado sería también como inventarse un nuevo tipo de minueto cambiando el orden de sus pasos (élevé derecha, plié, élevé izquierda, pas marché derecha, etc.). (N. del t.)

12

LOS ORÍGENES DEL LENGUAJE

No me cabe duda de que el lenguaje debe sus orígenes a la imitación y la modificación de diversos sonidos naturales y de los gritos distintivos del hombre, con ayuda de signos y gestos.

C

HARLES DARWIN, El origen del hombre

EL PROBLEMA DEL HUEVO Y LA GALLINA

Ahora que hemos visto cómo los memes en general, y las palabras en particular, crean el entorno en que crecen nuestras mentes, es momento de volver atrás y abordar los orígenes del lenguaje y de la cultura y ver qué podemos desentrañar acerca de cómo sucedió. En el capítulo 9 ya señalamos que el tema del origen del lenguaje es como el tema del origen de la vida; es probable que ambos sean sucesos únicos en este planeta que han dejado solo unas pocas y prometedoras pistas sobre su origen. La evolución del lenguaje se ha calificado como «el problema más difícil de la ciencia» (Christiansen y Kirby 2003, p. 1) y no cabe duda de que ha generado fuertes controversias desde poco después de que Darwin publicara El origen de las especies en 1859.

En realidad, uno de los memes más populares de la lingüística (un meme del que yo mismo he sido cómplice en su propagación, hasta ahora) es que el tema resultaba tan especulativo que la Société de Linguistique de

París prohibió cualquier discusión sobre la evolución del lenguaje en 1866, una política adoptada también por la Philological Society de Londres en 1872 (véase, p. ej., Corballis 2009). Durante décadas, esta historia ha formado parte del estandarte de muchos chomskianos que gustan de desestimar cualquier enfoque evolutivo del lenguaje como meros «cuentos de así fue», indignos de una evaluación científica. Sin embargo, resulta que la Société de Linguistique de París no defendía la ciencia sobria y formal, sino justo lo contrario: asumía una postura dogmática frente a los materialistas de la Société d'Anthropologie y se consagraba a fomentar puntos de vista monárquicos y católicos. Tal como he descubierto hace poco, gracias a Murray Smith:

La Société de Linguistique de París aún existe. Su sitio web tiene una página sobre su historia, que afirma que en 1866 se redactaron unos estatutos que especificaban (artículo 2) que «la sociedad no recibe correspondencia acerca del origen [no la «evolución»] del lenguaje ni de la creación de un lenguaje universal». Al crear esta regla, prosigue la historia, la sociedad pretendía diferenciarse de los círculos «positivistas y republicanos». La sociedad no estaba proscribiendo un tema controvertido, sino que estaba tomando partido sobre él. (correspondencia personal 2015)

Durante décadas, este rumor sobre la prohibición (demasiado bueno para ser cierto, tal como ha resultado) ha contribuido a rechazar los enfoques evolutivos en el lenguaje, pero avances recientes en los campos relevantes han incitado a intrépidos investigadores en lingüística, biología evolutiva, antropología, neurociencia y psicología a salir a la luz, con un amplio abanico de expectativas. Por muy fascinantes que puedan ser las diversas posturas, todo lo que necesitamos para nuestros propósitos es una visión general a vista de pájaro. Al considerar todas las posibilidades, deberemos mantenernos alerta por el peligro de la ceguera ante los gambitos (véase el capítulo 2), el bloqueo que asola a la imaginación adaptacionista. Como recalcamos antes, el primer ser vivo y reproductor pudo haber sido una torpe mezcolanza de partes ineficientes, que solo más tarde se fue mejorando gracias a la competición evolutiva, para llegar a ser las bacterias eficaces y austeras de hoy en día. ¿Cómo

pudo ser el ancestro de los bien diseñados lenguajes actuales? Seguramente era un conjunto de patrones conductuales ineficiente y de difícil aprendizaje, que pocas veces «funcionaba». ¿Qué condiciones tenían que darse para que valiera la pena invertir en estas primeras versiones? Puede que ni tan solo compensaran el esfuerzo de usarlas. Pueden haber sido hábitos parasitarios, infecciosos y difíciles de quitarse de encima. En consecuencia, tenemos que plantearnos la pregunta: ¿quiénes eran los principales beneficiarios, las palabras o los hablantes? Para muchos, ha parecido evidente que la respuesta tenía que ser «los hablantes», pero esto puede ser por no haber tenido nunca en cuenta el punto de vista de los memes. Tenemos que estar bien atentos en nuestra ruta tortuosa, en la que los gambitos nos acechan por todas partes, y si bien no hay duda de que actualmente el lenguaje ofrece un buen servicio a los anfitriones humanos, en sus primeros días puede que fuera más una carga que una bendición. Sea como fuere que se iniciara el lenguaje, a continuación presento una lista conjetural de las funciones a las que, con el tiempo, todos los lenguajes prestaron servicio:

Utilidad comunicativa. Es decir, la capacidad del lenguaje para ordenar, pedir, informar, preguntar, instruir, insultar, inspirar, intimidar, tranquilizar, seducir, divertir, entretener.

Productividad. Es decir, la capacidad del lenguaje de generar un Vasto (véase el capítulo 6, nota 14, p. 382) número de significados (frases, expresiones) diferentes formados a partir de una reserva finita de elementos léxicos. Formalmente, no hay límite a la cantidad de frases gramaticales del español (no hay ninguna regla que diga, por ejemplo, que ninguna frase pueda tener más de n palabras), pero incluso si tuviéramos que restringirnos a frases de no más de veinte palabras, por decir algo, el número de frases gramaticales con sentido, fácilmente comprensibles por adultos normales tras oírlas una sola vez, seguiría siendo Vasto (la frase previa casi con toda seguridad nunca antes ha sido expresada por nadie, y tiene 66 palabras, pero no ha sido tan difícil, ¿verdad?).

Digitalidad. Es decir, la capacidad de los receptores y transmisores del lenguaje de «corregir a la norma», tal como hemos visto, con lo que se filtra buena parte del ruido de la señal, incluso sin comprensión.

Referenciación remota. Es decir, la capacidad del lenguaje de referirse a cosas que no están presentes en el entorno de los comunicadores, fuera del campo de visión, en el pasado o en un entorno imaginario o hipotético.

Facilidad de adquisición. Es decir, la notable celeridad con que los niños captan los lenguajes hablados o gestuales (en comparación con la lectura, la escritura o la aritmética, por ejemplo).

Parecería que otros mamíferos sociales (como simios, lobos, elefantes, ballenas o murciélagos) podrían emplear el lenguaje de forma espectacular, si dispusieran de él, y es cierto que algunos poseen talentos comunicativos bien conocidos, pero ninguna otra especie tiene una facultad remotamente parecida al lenguaje humano.[1] De algún modo, nuestros ancestros sin lenguaje se toparon con una vía poco frecuentada que los llevaba hacia ese tesoro y la siguieron a tientas, encontrando algunas ventajas por el camino, o por lo menos sin rezagarse demasiado, hasta que les tocó el gordo. ¿Cómo y por qué? Y ¿cómo es que otras especies no han hallado el mismo Buen Truco?

No se me ocurre nadie que niegue que un precursor necesario del lenguaje ha tenido que ser algún tipo de transmisión cultural prelingüística respaldada gracias a algún ajuste genético. Por ejemplo, en muchas especies sociales sin lenguaje encontramos el instinto de cooperar con la propia familia ampliada de individuos conespecíficos, como sucede con los perritos de las praderas, los elefantes y las ballenas; si a esto añadimos un apremiante hábito instintivo de imitación (de los padres, de los ancianos, de los compañeros), podría ser suficiente para crear un entorno de apoyo para la evolución cultural desbocada que acabó apareciendo en nosotros, a caballo del lenguaje. Dicho de otro modo, las palabras pueden ser los mejores memes, pero no fueron los primeros. Este tipo de instintos no progresarían ni se desarrollarían sin alguna innovación ambiental o conductual que subiera el listón para los futuros individuos de la población, creando una presión selectiva hacia una mayor propensión a cooperar (cooperar con más rapidez, con más facilidad, con más atención). Igualmente, el instinto de copiar a los propios ancianos podría ser un intensificador de la aptitud en muchas situaciones, aunque no en todas. ¿Revisó la evolución genética una

estirpe de homínidos y les otorgó instintos que, de un modo u otro, les permitió compartir innovaciones culturales más que otras estirpes? ¿Qué nueva presión selectiva, restringida a un único grupo, podría poner en marcha tal proceso? ¿Evolucionó la cooperación grupal antes que el lenguaje? Los chimpancés a menudo se dedican a una caza más o menos cooperativa de colobos. ¿Acaso un nuevo reto del entorno obligó a un grupo de homínidos (nuestros antecesores) a confiar más en la cooperación? ¿Hizo esto aparecer nuevos hábitos de atención? (Tomasello 2014). ¿Qué beneficio obtuvieron los hijos de nuestros ancestros tan interesados en las vocalizaciones de su grupo y tan ansiosos por imitarlas?

Richerson y Boyd, en Not by Genes Alone (2004) casi no dicen nada sobre el lenguaje, deliberadamente. A medio libro afirman «hasta ahora no hemos dicho nada sobre el lenguaje y la razón es simple: los paleoantropólogos no tienen la menor idea de cuándo apareció el lenguaje humano» (pp. 143-144). Hay un intervalo de dos millones de años para la controversia, con algunos detalles neuroanatómicos (deducidos de cráneos fosilizados de homininos) que sugieren que los elementos del lenguaje han estado presentes desde hace millones de años, mientras que otras pruebas dan a entender que incluso hace 50.000 años el lenguaje era rudimentario, como mucho.

¿Es necesario el lenguaje para transmitir el conocimiento de la talla lítica con la que construir herramientas? ¿Para comerciar (materiales, comida, utensilios)? ¿Para hacer fuego? ¿Podemos imaginar a los jóvenes homininos adquiriendo el autocontrol y la previsión necesarias para cuidar el fuego de forma eficaz sin instrucciones verbales (y después autoadvertencias)? Según Sue Savage-Rumbaugh (correspondencia personal 1998), a los bonobos en cautividad les gusta sentarse alrededor de una hoguera, pero ¿podríamos entrenarlos para que recogieran madera y que cuidaran un fuego indefinidamente, conservando el combustible y evitando su descontrol? Sea cual sea la respuesta, no nos aportaría demasiada luz acerca de si nuestros ancestros podían enseñarse, unos a otros, a cuidar del fuego antes de que tuvieran por lo menos una versión muy basta del lenguaje. ¿Por qué no? Porque los bonobos han evolucionado, separándose de nuestros comunes ancestros, tanto como nosotros (unos seis millones de años) y hay todo un mundo de diferencias, inestudiables e indeterminables, entre la perspectiva y los talentos de los entrenadores humanos formados y equipados con lenguaje

que pueden imponer programas de formación a los bonobos cautivos y la perspectiva de un hominino adulto criando a su descendencia. Tal como señalé una vez (1978), «los osos pueden ir en bicicleta, un hecho sorprendente con una escurridiza implicación teórica». Las pinturas rupestres de Lascaux (de una antigüedad de entre 20.000 y 30.000 años), ¿pueden haber sido realizadas por Homo sapiens artistas sin lenguaje? (véase Humphrey 1998). Estas preguntas clave, y muchas otras, permanecen sin respuesta, y quizá no la puedan tener.

Una ventaja de aplazar cualquier hipótesis sobre cuándo entró en escena el lenguaje funcional es que permite a Richerson y Boyd explorar modelos minimalistas por lo que respecta a un aspecto importante: presuponen una ausencia casi total de comprensión, porque presuponen ausencia de comunicación (cuando mis alumnos leen su libro suele ayudarles preguntarse: «¿que sería necesario para que la transmisión cultural prosperara entre las tortugas, entre las gaviotas o entre las ovejas?»; es una manera de evitar que prematuras suposiciones de comprensión por parte de los actores implicados nublen la imaginación). Por ejemplo, Richerson y Boyd exploran modelos en que se da una transmisión sesgada, y este sesgo puede ser cualquier cosa que haga que unos memes se transmitan con más probabilidad que otros.[2] La transmisión sesgada cuenta como fuerza selectiva, tanto si se trata de un sesgo bueno, inteligente o comprendido como si no. Un sesgo propenso a ser más valioso que «copia cualquier cosa que se mueva» o «copia al primer adulto que veas» es «copia a la mayoría» (sesgo conformista), «copia a los más capaces» o «copia a los más prestigiosos». Aunque, siempre que B copia a A, hay información que pasa de A a B, esto no se trata siempre de comunicación, no más que lo es que A pase un resfriado a B (aunque hablemos de enfermedades transmisibles).

Con estas condiciones minimalistas, ¿cuándo un hábito de copia (básicamente a ciegas) funciona mejor que enfrascarse en un aprendizaje a base de pruebas y errores? ¿Mejor para quién? ¿Cui bono? ¿Mejor para los copiadores individuales, para la población que adopta los hábitos de copiar o para los propios memes? Richerson y Boyd son conscientes de que la cultura maladaptativa también se puede propagar, pero sostienen que hasta que la cultura adaptativa (cultura que mejora la aptitud genética de los que la poseen) logra establecerse, los hábitos y órganos mantenidos genéticamente de los que depende la cultura se seleccionan negativamente con rapidez. El paralelismo entre memes y virus clarifica

el razonamiento: los virus no se pueden reproducir por sí mismos, ya que dependen de poder controlar la fiable maquinaria de copia de los núcleos celulares, una maquinaria producto de miles de millones de años de I+D. Los memes, provechosos o no, por encima de todo deben copiarse a sí mismos, de modo que por lo menos toda esa infraestructura formada genéticamente (propensiones a prestar atención a los demás y a copiar algunas de las maneras de hacer percibidas) es la única base sobre la cual los memes pueden arraigar y prosperar.

Se necesita gran cantidad de trabajo de diseño para crear y optimizar un medio de intercambio de información fácil de usar y, además, «alguien tiene que pagarlo», pero una vez que ya hay en marcha un sistema de copia rudimentario, intrusos egoístas pueden apropiarse de él. Quizá solo seamos monos con cerebros manipulados por memes, tal como somos manipulados por los virus del resfriado. En lugar de mirar solo a los prerrequisitos en competencias que nuestros ancestros tenían que tener para que el lenguaje se pusiera en marcha, quizá deberíamos tener también en cuenta vulnerabilidades inusuales que podrían hacer de nuestros ancestros los anfitriones ideales de hábitos (memes) infecciosos, pero no virulentos, que nos permitían vivir lo bastante y con la movilidad suficiente para que esos memes se propagaran por nuestras poblaciones.

El resultado de la evolución cultural, hasta la fecha, ha sido un aumento espectacular de la población humana, con cientos de miles de memes en cada rincón del planeta, incluyendo muchos entornos que no hubieran podido ser ocupados por Homo sapiens sin su ayuda. Pero tal como la hormiga que sube por una brizna de hierba es la forma que tienen los trematodos de llegar al intestino de una vaca o de una oveja, quizá deberíamos considerar a los astronautas que han llegado a la Luna como la forma que tienen los memes de llegar a la siguiente generación de fanáticos de la ciencia.[3]

El punto de vista de los memes permite una pequeña corrección a los diversos modelos de herencia dual o de doble vía de la relación entre la evolución genética y la cultural (Cavalli-Sforza y Feldman 1981; Boyd y Richerson 1985, entre otros). Las adaptaciones (potenciadores de aptitud) pueden transmitirse genéticamente o culturalmente. La vía de la información genética se ha optimizado a lo largo de miles de millones de años, una maravilla de ingeniería de asombrosa complejidad, con copiadoras de ADN, máquinas de edición y sistemas para encargarse de

los parásitos genómicos (ADN vagabundo que se instala en los genomas, donde el daño que provocan queda más o menos controlado y minimizado).[4] La vía cultural, en un periodo mucho más breve, también ha generado una multitud de características de diseño para facilitar la transmisión fiable de información y buena parte de la I+D que ha creado y ajustado estas características ha sido la evolución de los propios memes, para «encajar en el cerebro», mientras que los ajustes genéticos del cerebro han llegado a continuación; un proceso coevolutivo en que, en general, la «investigación» corre a cargo de los memes y el «desarrollo» posterior, a cargo de los genes. Es decir, las innovaciones en los memes que los hicieron reproductores más eficaces en cerebros que aún no estaban lo bastante bien diseñados para lidiar con ellos pudieron ofrecer las primeras «pruebas de viabilidad» que respaldarían los ajustes genéticos, más caros y prolongados, en la estructura del cerebro que, a su vez, mejorarían las condiciones de trabajo de los memes y de sus anfitriones.

Este patrón básico se ha repetido centenares de veces desde el inicio de la era de la informática, con innovaciones en el software que llevan posteriormente a rediseños del hardware una vez que las versiones de ese software han demostrado funcionar. Si comparamos los chips actuales con sus antecesores de hace cincuenta años, veremos muchas innovaciones que, originalmente, fueron diseñadas como sistemas de software, como simulaciones de nuevas computadoras ejecutándose en el hardware entonces existente. Una vez demostradas sus virtudes y eliminados o minimizados sus defectos, este software podía servir de especificaciones para construir nuevos chips procesadores, versiones mucho más rápidas de las simulaciones. A medida que surgían nuevos entornos para las computadoras, se exploraron competencias conductuales primero en forma de programas ejecutándose en computadoras genéricas, en las cuales la revisión es rápida y barata, y luego se probaron extensamente en la práctica frente a los competidores; los mejores diseños acabaron incorporándose en hardware específico. Por ejemplo, los teléfonos inteligentes actuales, además de capas y capas de software ejecutándose en software ejecutándose en software, disponen de hardware específico para gráficos y para reconocimiento y síntesis de voz en sus microprocesadores, descendientes de sistemas de software que fueron los primeros en explorar el espacio de diseño.

Toda esta I+D informática ha sido un proceso de diseño inteligente de arriba abajo, por supuesto, con un profundo análisis de los diversos aspectos del problema y de las cuestiones de acústica, óptica y demás campos de la física, y bajo la guía de aplicaciones explícitas de análisis de rentabilidad, pero aun así también ha destapado muchas de las mismas rutas hacia un buen diseño ya localizadas a ciegas por el diseño darwiniano de abajo arriba a lo largo de periodos más prolongados. Por ejemplo, vuestro teléfono inteligente tiene un hardware específico para el procesamiento de la voz, pero no para hablar inglés o chino; estas competencias se pueden añadir con software en cuanto el teléfono haya llegado a su entorno local. ¿Por qué? Por la misma razón (la misma motivación flotante) por la que el cerebro de un bebé es idiomáticamente neutro: la versatilidad amplía el «mercado» disponible para el diseño. Esta situación puede cambiar: si se acabara fijando un único lenguaje humano para todo el mundo (un proceso augurado por los cientos de lenguas que se extinguen cada año), los cerebros de los bebés que, casualmente, tuvieran un sesgo incorporado para aprender ese lenguaje, recibirían una selección positiva y, tras varias generaciones, se llegaría a una eventual pérdida de versatilidad lingüística en el hardware humano fundamental. Esto sería un claro ejemplo del efecto Baldwin, que muestra cómo se reduce la variedad y la versatilidad genéticas llevando un comportamiento (o una opción de desarrollo) hacia un corsé de «prácticas recomendadas» controlado genéticamente y convirtiendo opciones en comportamientos obligados, tal como vimos en el capítulo 9.

Una vez construida la vía cultural, en cualquier colaboración estable entre rasgos meméticos y rasgos cerebrales, entre software y hardware, pueden medrar parásitos culturales (variantes culturales solitarias, en la terminología de Boyd y Richerson, o memes parasitarios, en la de Dawkins), a pesar de las posibles contramedidas, en una carrera armamentística como cualquier otra de las que se producen en la evolución.

No todos los entornos son un terreno abonado para la evolución de la evolución cultural. Si pensamos en los copiadores como «gorrones» y en los aprendices como productores de información (Richerson y Boyd, p. 112, citando a Kameda y Nakanishi 2002), podemos ver que tiene que haber un término medio, con la posibilidad de un equilibrio en el que algunos innovadores carguen con los costes de la I+D (no de la vía de la información, sino de la invención o descubrimiento de lo que sea que

produzcan) y el resto se sirvan de la información a un coste más bajo.[5] El hecho de que modelos simples muestren este equilibrio dinámico sugiere que podría existir en una población sin lenguaje pero con la suficiente variación en la curiosidad (lo que favorecería la I+D, por más que sea cara) y la conformidad (lo que favorecería la copia, aunque fuera de información anticuada).[6] Sin embargo, esto depende de la variabilidad del entorno: si es demasiado predecible o demasiado impredecible, la transmisión cultural (la copia) no puede asegurar su posición. Esto nos ofrece un ejemplo de una posible condición liminar: la transmisión cultural no aparecerá excepto en un entorno de condiciones favorables (un entorno del tipo «ricitos de oro») que no sea ni muy caliente (caótico) ni muy frío (invariable) y que se mantenga durante el tiempo suficiente para que la evolución tenga la oportunidad de crear nuevos hábitos y fijarlos en la población. Si estos largos periodos son poco frecuentes, es muy posible que las poblaciones que hubieran estado bien dotadas para aprovecharse de la transmisión cultural no lleguen a dar los pasos necesarios.

La cultura ha sido un Buen Truco con un éxito sin precedentes para Homo sapiens. Cualquier teoría del origen de la cultura que no plantee que hay por lo menos un umbral, una barrera para su adopción y que tal umbral ha sido superado de algún modo, se enfrenta a la objeción de que si adquirir la transmisión cultural fuera tan fácil, muchas especies (mamíferos, aves, peces) ya la tendrían. Un entorno con un historial de la cantidad adecuada de variación, ni muy monótono ni muy accidentado, es un posible umbral, pero se han propuesto otros. Consideremos el bipedismo, la marcha erguida sobre dos piernas, un rasgo favorito desde que Platón describiera a los humanos como bípedos sin plumas. Entre los homínidos, solo los homininos presentan esta característica, que sin duda liberó las manos de nuestros ancestros para poder elaborar artefactos y transportarlos de un sitio a otro, así como las materias primas necesarias. Pero la evolución no es previsora, de modo que la motivación flotante original de esta excelente innovación de diseño tiene que haber sido alguna otra cosa. He aquí otro dilema del huevo y la gallina: ¿creó la elaboración rudimentaria de herramientas (como la de los chimpancés) una presión selectiva para la capacidad de transportar materias primas o artefactos acabados a largas distancias? ¿O la marcha erguida apareció por otras razones y abrió así el espacio de diseño a una elaboración de herramientas eficaz? Las hipótesis son muchas. La hipótesis de la sabana plantea que cuando un clima más seco expulsó a nuestros ancestros de

los árboles, obligándoles a ocupar las praderas, la locomoción erguida se vio favorecida, ya que permitía ver más lejos por encima del pastizal (o minimizaba la exposición al Sol o minimizaba la energía consumida en la locomoción), mientras que las diversas variantes de la hipótesis del simio acuático (A. Hardy 1960; Morgan 1982, 1997) proponen que la innovación ecológica fue la recolección de moluscos en aguas poco profundas. O quizá las plantas acuáticas (Wrangham et al. 2009) se convirtieron en un alimento «de último recurso» en tiempos difíciles, forzando a los individuos a entrar en aguas cada vez más profundas, a aguantar la respiración y acaso a otros cambios fisiológicos. Todas estas hipótesis son controvertidas y es posible que lo sigan siendo durante algún tiempo; en cualquier caso, ¿podría ser que el bipedismo y su resultante conjunto de competencias activadas abrieran las puertas al lenguaje y la cultura?

Otro umbral que se ha propuesto es la inteligencia social (Jolly 1966; Humphrey 1976): la competencia para interpretar a los otros como sistemas intencionales cuyas acciones se pueden anticipar observando lo que estos otros observan y determinando lo que quieren (comer, huir, atacar, aparearse, estar a solas). Esta competencia a menudo se llama teoría de la mente (TdM o TOM, Theory of Mind), una denominación mal escogida, porque nos

sugiere que los que tienen esta competencia deben ser unos teóricos con

comprensión, astutos recolectores de pruebas y planteadores de hipótesis, en lugar de agentes anticipadores instintivos, bendecidos con un talento interpretativo que no tienen por qué entender. En cualquier caso, podría ser que esta perspectiva fuera necesaria para la transmisión de información cultural compleja, tal como el bipedismo parece necesario para una cultura material compleja formada por herramientas, armas, viviendas, recipientes, vestidos y barcas.

Según Michael Tomasello (2014), otro investigador destacado en el campo de la evolución de la cognición humana, esta perspectiva, aparecida inicialmente para mejorar la competición (entre conespecíficos y en la caza) podría evolucionar hacia un instinto de cooperación, cuyas manifestaciones rudimentarias se pueden ver actualmente en la caza de los chimpancés, por ejemplo, y luego evolucionar hacia «procesos aún más sofisticados de intencionalidad conjunta, con objetivos comunes y

atención común, creada para la coordinación social» (p. 34). Tomasello afirma que

nuevas presiones ecológicas (p. ej. la desaparición de alimentos que se podían obtener individualmente y luego una mayor población y competición por parte de otros grupos) actuaron directamente sobre la organización y la interacción social humana y llevaron a la evolución de formas de vida humana más cooperativas (como la colaboración para la búsqueda de comida y luego, organización cultural para la coordinación y la defensa grupal). (p. 125)

Tomasello considera el lenguaje como «el remate de la cognición y el pensamiento exclusivamente humanos, no su fundamento» (2014, p. 127), pero aunque coincido con su énfasis en la cantidad y complejidad de la evolución cultural y genética prelingüística, sin la cual no habría lenguaje, cuando apareció el lenguaje, este permitió una evolución cultural acumulativa (y, de hecho, una evolución genética como respuesta) en un proceso desbocado de acumulación cada vez más rápida y eficaz de nuevos diseños y descubrimientos. Puede que el lenguaje no sea el fundamento, pero yo tampoco lo consideraría el remate final; más bien diría que es la plataforma de lanzamiento de la cognición y el pensamiento humanos.

Laland, Odling-Smee y Feldman (2000) introdujeron el concepto de construcción de nicho, la idea de que los organismos no solo responden al entorno selectivo en el que nacen, sino que, además, sus actividades pueden modificar con rapidez las características de ese entorno, creando así nuevas presiones selectivas y aliviando otras. De este modo, el nicho en que medran sus descendientes puede ser diferente, en aspectos importantes, del nicho con el que lidiaron sus progenitores. La construcción de nicho no es solo un efecto de las presiones selectivas de la selección natural, también es una causa relevante, e incluso desestabilizante, de nuevas presiones selectivas, una grúa con una considerable capacidad de carga en el espacio de diseño. No cabe duda de que nuestra especie se ha implicado a fondo en la construcción de nichos.

Steven Pinker (2003, 2010) llama a nuestro mundo nicho cognitivo, remarcando que es un producto de la comprensión humana. Boyd, Richerson y Henrich (2011) no están de acuerdo con Pinker y proponen que sería mejor denominarlo nicho cultural, una plataforma de competencias en la cual puede crecer la comprensión. Como veremos, la I+D que ha construido el nicho en que vivimos hoy en día es una mezcla cambiante de procesos darwinianos de abajo arriba y diseño inteligente de arriba abajo. Sin duda, nuestro nicho no se parece al de ninguna otra especie; apenas incluye presas ni depredadores (a no ser que seas un pescador o un surfista en aguas infestadas de tiburones), los hábitats están formados por casi nada más que artefactos y plantas y animales domesticados y los papeles sociales, la riqueza, la reputación, la experiencia y el estilo o la moda (al vestirse, al hablar, al cantar, al bailar, al jugar) han sustituido casi por completo a unos músculos más potentes, a una mayor velocidad al correr y a una visión más aguda como ventajas que aportan aptitud genética.

El papel que ha tenido el lenguaje en permitir y mejorar esta transformación es tan obvio que hay una cierta tradición de crítica correctora, que insiste en que por lo menos las primitivas versiones de la agricultura, la pesca, el vestido, la religión, la decoración, la preparación de comida y otros elementos habituales de la cultura podían prosperar y transmitirse sin lenguaje. La preparación ritual de los cadáveres para su entierro, por ejemplo, sugiere claramente algo parecido a una creencia en la vida después de la muerte, pero resulta difícil ver como algo así podría compartirse sin expresión verbal. Después de todo, las formas lingüística y no lingüística de transmitir información costosamente obtenida continúan coexistiendo, de modo que, como siempre, podemos suponer que la interacción humana a lo largo de los milenios condujo a la adopción gradual de formas cada vez más eficaces y sistemáticas (memes), incluyendo formas de adquirir memes, como el paso de la burda imitación al aprendizaje (Sterelny 2012). No hay duda de que algunas de estas transmisiones de memes exigían la atención conjunta, otras exigían una guía (proto)lingüística y otras exigían instrucciones plenamente lingüísticas, como mantras mnemónicos y otros recursos.

Podemos prestar atención al siguiente ejercicio de la imaginación: preguntémonos cómo (en detalle) un animal progenitor (un perro, un lobo, un delfín, un hominino sin lenguaje) podría, sin lenguaje, transmitir a sus hijos parte de su experiencia tan esforzadamente lograda. Por

ejemplo, supongamos que un lobo ha aprendido a las malas que hay que evitar los puercoespines, en lugar de abalanzarse sobre ellos. ¿Cómo pasará el beneficio de su experiencia a sus crías? Quizá llevándolas hasta un puercoespín cercano y entonces hacerles señales de «¡evitadlo!» que capten su atención, de un modo u otro. Esto exige una atención conjunta y la confianza u obediencia de la descendencia (que es lo que hacen los padres humanos con sus niños prelingüísticos cuando les advierten de mantenerse alejados de las estufas, con éxito variable). Pero en ausencia de un ejemplo que ocupe la atención conjunta del instructor y el aprendiz, no tiene ningún otro recurso. La capacidad del lenguaje para centrar la atención en cosas y situaciones ausentes es una mejora formidable.

En su último libro, Derek Bickerton (2014) considera que esta capacidad de «referenciación remota» es la innovación clave del lenguaje, pero plantea una mirada ligeramente diferente a la pregunta que estamos considerando, a la que alude en su título: More Than Nature Needs: Language, Mind, and Evolution. «¿Cómo adquirió la especie humana una mente que parece mucho más potente que lo que los humanos podían necesitar para sobrevivir?» (p. 1). Según Bickerton, «la laguna cognitiva entre animales humanos y no humanos es el talón de Aquiles de la evolución» (p. 5) y, en su opinión, no se puede explicar mediante ninguna descripción de la selección natural directa de comportamientos comunicativos cada vez más potentes que, finalmente, acabaron dando lugar al lenguaje. Por ejemplo, los animales sociales necesitan llevar un control de los cambios de rango y de quién hizo qué a quién, pero estos chismes estratégicamente valiosos no se podían transmitir

hasta que el lenguaje hubiera llegado a un considerable grado de desarrollo. Dicho de otro modo, en las primeras fases del lenguaje, este no puede haber expresado ningún chisme que tuviera el más mínimo interés y, dado que tuvo que haber algún otro motivo que impulsara estas primeras fases, el chismorreo como sustituto preparativo no pudo haber constituido la presión selectiva como base del origen del lenguaje. (p. 63)

Esto recuerda los habituales argumentos creacionistas de que hasta que el ojo, las alas o el flagelo de las bacterias no está totalmente diseñado, es inútil y, por tanto, no pudo haberse mantenido generación tras generación; el típico reto al evolucionismo basado en «no puedes ir de aquí hasta allí». Pero Bickerton no es ningún creacionista y tras mostrar lo que él cree que son fallos en los intentos de otros, ofrece su solución a la «paradoja de la cognición» (p. 79):

Si descartamos los procesos selectivos normales y las versiones mágicas de la evolución, ¿qué nos queda? La respuesta es que solo nos quedan dos cosas: palabras y sus consecuencias neurales. La invención (porque solo pudo ser una invención, si bien no consciente, ni tan solo intencionada) de unidades simbólicas tuvo unas consecuencias cruciales para el cerebro. (p. 91)

Me gusta la ironía de este fragmento, con su dicotomía entre evolución e invención, y su aserción de un salto de creatividad prometeico e inconsciente, porque durante más de una década desde que nos conocimos, Derek ha disfrutado retándome a presentarle un solo ejemplo de algún fenómeno social o cultural que solo se pudiese explicar mediante los memes. Ahora él mismo ha respondido a su desafío no con uno, sino con dos ejemplos: el lenguaje humano y la cognición; Derek no expresa la respuesta en términos de memes (que no aparecen ni en el índice alfabético de su libro), pero vale la pena citar algunos pasajes muy reveladores:

¿Cómo reaccionaría el cerebro a tener palabras en él? (p. 100)

[…] las reacciones más probables del cerebro a su colonización por parte de las palabras. (p. 108)

Expresar palabras es un proceso autocatalítico: cuanto más las usas, más rápido te vuelves hasta llegar a un punto en que puedes formar expresiones mentalmente antes de articularlas. (p. 123)

Un poco antes, en su libro, hay un fragmento que muestra el punto exacto en qué Derek perdió el tren:

Tengamos presente que el comportamiento (proto)lingüístico tuvo que merecer la pena desde las primeras expresiones; de lo contrario, su infraestructura genética no se hubiera fijado nunca. Así pues, consideremos, teniendo en cuenta nuestra presión selectiva favorita, cuáles pudieron haber sido estas primeras expresiones y si pudieron haber sido lo bastante útiles para conservarse e incorporarse. (p. 63)

¡O lo bastante infecciosas! Derek ignora la posibilidad de que una vez instalado un hábito de copia sin comprensión, impulsado por el tipo de eficacia comunicativa que podía proporcionar una señalización simple («animal»), las presiones selectivas sobre los memes pueden haberse hecho cargo, independientemente de la utilidad (para la aptitud humana) de los memes que se están elaborando y propagando.[7] Como especie sinantrópica, tenían que captar la atención y ser inolvidables, pero no particularmente útiles, por lo menos al inicio.

A partir de los trabajos de Tomasello, Bickerton cree que el paso de la cooperación conjunta al lenguaje plenamente desarrollado tiene que haber pasado por un carroñeo confrontacional: el clima cambió, lo que obligó a nuestros antecesores a modificar su dieta y volverse carroñeros de la sabana, pero tenían que eludir a los depredadores y a otros carroñeros, por lo que necesitaban grupos grandes equipados con afiladas lascas de piedra y probablemente lanzas. La búsqueda de piezas para carroñear en grandes grupos habría sido ineficiente: es mejor enviar exploradores, en solitario o en parejas, en muchas direcciones para que regresen al poblado a informar (como las abejas) cada vez que encuentren una pieza aprovechable. Al igual que las abejas, lo mejor es que proporcionen información acerca de dónde estaba la fuente de

comida y cuán valiosa era. Según Bickerton, este fue el nacimiento de la referenciación remota, el poder semántico (que no se halla en las simples llamadas de alarma o de apareamiento) que puede centrar la atención en algo que no se encuentra en el campo visual ni en el campo auditivo. Son las abejas la especie a sopesar, no los bonobos.

Tiene que haber habido alguna poco frecuente secuencia de umbrales o cuellos de botella que nuestros ancestros atravesaron a tientas. Sin duda, no reflexionaron inteligentemente: «¡Diseñemos un lenguaje y así podremos coordinar nuestras actividades y controlar todo lo que nos rodea!». Y también parece obvio que no se fueron haciendo cada vez más listos que sus parientes homininos gracias a unos «mejores genes de la inteligencia» hasta que fueron lo bastante listos como para transformar sus gruñidos y aullidos de hombre de las cavernas en un lenguaje gramatical. Las mutaciones sin presiones selectivas que las favorezcan desaparecen en pocas generaciones. De algún modo, nuestros ancestros se encontraron en una situación en la que había la poco frecuente oportunidad de obtener una gran recompensa. ¿Qué es lo que evitó que otros homínidos, sometidos a los mismos retos, no estuvieran a la altura de las circunstancias? Richerson y Boyd (2004) llaman a este problema el dilema adaptacionista (p. 100), y señalan:

La incerteza acerca de lo que pasaba en el Pleistoceno temprano es cada vez más clara, y saber qué es lo que no sabes es tan importante como saber qué es lo que sabes. (p. 142)

El origen último del lenguaje y el origen de la vida (u orígenes) son problemas aún sin resolver, pero en ambos casos no escasean las historias plausibles «de así fue», a la espera de una refutación o de convertirse en hipótesis confirmadas, una gran reserva de donde elegir para los investigadores futuros.

LOS TORTUOSOS CAMINOS HACIA EL LENGUAJE HUMANO

Después de haber más o menos asegurado una cabeza de puente, la evolución posterior de los fenómenos protolingüísticos también aporta abundantes oportunidades para la controversia y la mezcolanza:

Inicialmente, había un protolenguaje de breves expresiones y articulaciones, como las llamadas de aviso de los cercopitecos verdes,[8] sin productividad y sin distinción alguna entre expresiones imperativas y enunciativas que pudieran diferenciar «¡sube al árbol!» de «se acerca un leopardo» (Bickerton 2009, 2014; véase también Millikan 2004, sobre las representaciones «pushmi-pullyu»).******** Sugeridas por tipos de sucesos cruciales en la vida de los homininos, estas señales serían reacciones adecuadas y reconocibles a ofrecimientos importantes y, por tanto, serían ellas mismas ofrecimientos, elementos del Umwelt de los homininos, sin que se necesite más análisis semántico por parte de los beneficiarios que el necesario para las llamadas de alarma de otras especies.

Quizá en primer lugar surgió un lenguaje gestual, parecido a las lenguas de signos actuales, con vocalizaciones para captar la atención y remarcar aspectos diversos (Hewes 1973; Corballis 2003, 2009). Hablar sin hacer gestos resulta difícil para muchas personas y podría ser que la vocalización y la gestualización hayan intercambiado sus papeles, de modo que ahora los gestos funcionan como complemento y ayuda, cuando antes este era el papel de las vocalizaciones. Los movimientos manuales vestigiales que muchos no podemos evitar al hablar podrían ser restos fósiles de los lenguajes originales.

Acaso se produjo una carrera armamentística auditiva del tipo «cola de pavo», con los machos homininos rivalizando para mostrar su talento en vocalizaciones musicales, quizá con improvisaciones, como las exhibiciones de los ruiseñores y otros pájaros cantores. Las sílabas no tendrían que tener ningún significado ni función más allá de su atractivo, y es aquí donde se podía favorecer la productividad, al igual que en la música, como recurso para proporcionar una mezcla de novedad (agradable) y familiaridad (agradable), tal como suele suceder con la música. Laralá laralá, oeoeee oeoeeee, parapam pam pam, ninonino… Las hembras desarrollarían su propio don de parloteo sin sentido como

subproducto de las habilidades cruciales que tenían que juzgar en la competición entre machos. Igual que los humanos, muchas aves tienen que aprender sus cantos, que son claras señales honestas de superioridad en la competición del apareamiento, y se dan tentadores paralelismos neuroanatómicos entre los cerebros de los pájaros cantores y los cerebros humanos (Fitch, Huber y Bugnyar 2010), pero tal como subraya Hurford (2014), la competencia lingüística se desarrolla antes de la pubertad y «atraer parejas sexuales antes de poder reproducirse sería un despilfarro de recursos y un método potencialmente perjudicial» (p. 172). Sin embargo, este argumento no es definitivo, y la selección sexual (véase, p. ej., Miller 2000) puede haber tenido un papel en algún momento en la formación de la virtuosidad lingüística.

Un análisis claro y exhaustivo de estas opciones, y otras más, aparece en The Origins of Language: A Slim Guide, de James Hurford (2014), un trabajo que se diferencia de otros por su reconocimiento (aunque no explícito) de que estas preguntas no son solo preguntas cómo es que, también son preguntas para qué.[9] Por ejemplo, el lenguaje tiene dos sistemas diferentes de composición, el fonotáctico (que determina qué fonemas pueden seguir a otros, independientemente del significado, lo que excluye nfik y sgopn como palabras del español, por ejemplo) y el morfosintáctico (que determina el orden de las palabras y el uso de prefijos y sufijos para crear significados a partir de significados). ¿Por qué hay dos niveles de composición, uno semántico y el otro no? La estructura composicional fonotáctica está determinada básicamente por las limitaciones del control vocal, la audición y la memoria:

Nuestras lenguas solo pueden lograr un conjunto limitado de sonidos del habla independientes, mientras que nuestros oídos solo pueden detectar diferencias acústicas hasta un cierto grado de sutileza.

Así, las consideraciones de economía física y de eficiencia dictan la solución al problema composicional en el ámbito fonológico, pero para empezar, ¿qué es lo que empuja la búsqueda de composicionalidad?

Si podemos situar estos sonidos en secuencias memorizadas, y tenemos la memoria suficiente para almacenar miles de tales secuencias (es decir, palabras), tendremos una solución eficiente para la tarea de expresar grandes cantidades de significados, dada una sintaxis semánticamente composicional. (p. 127)

La productividad de las lenguas está «motivada por» la utilidad de poder comunicar muchas cosas acerca del mundo (p. 128). Esta es una de las maneras en que Hurford alude a las motivaciones flotantes que responden a las preguntas para qué sobre la evolución del lenguaje. El propósito o raison d'être de la productividad es amplificar el poder expresivo de un sistema de comunicación sin tener que multiplicar los elementos hasta cantidades poco manejables. Sin embargo, conviene que nos detengamos a considerar la observación de Bickerton de que las primeras expresiones del protolenguaje no podían comunicar «una vasta cantidad de significados», de modo que esta nueva tarea no habría sido un paso evidente a intentar por parte de los usuarios de ese protolenguaje. Podemos comparar esta tarea con la de cómo obtener unas frutas visibles y apetecibles de unas ramas que están fuera de nuestro alcance: sabemos que la fruta es buena y, por tanto, la motivación para enfrascarse en agotadores intentos de prueba y error está claramente disponible. Pero está claro que una escalera improvisada es un resultado más probable que una gramática improvisada: ¿dónde está la seductora recompensa en los primeros días del lenguaje? Ningún inventor del lenguaje tuvo el sueño de crear un sistema útil de comunicación (tanto como una célula pueda soñar en convertirse en dos células), pero retrospectivamente la motivación está bien clara. ¿Qué proceso podría destapar esa motivación? Hurford no se compromete con ninguna respuesta a la pregunta cómo es que, pero en el tema de la pronunciación tiene un fragmento relevante:

En función de la edad y de la personalidad, la gente acaba hablando como las personas que tiene alrededor, a menudo sin ningún esfuerzo consciente. La evolución de los sistemas vocálicos es un caso de «autoorganización». Un sistema no evoluciona gracias a una

planificación deliberada, sino mediante la acumulación, a lo largo del tiempo, de un sinnúmero de pequeños ajustes por parte de individuos que responden a presiones inmediatas. (p. 153)

Esta aproximación hacia la sistematicidad y la productividad parece ser un proceso guiado por dos motivaciones que se refuerzan mutuamente. A los memes audibles, con significado o no, «les convenía» diferenciarse de la competencia, pero también aprovecharse de los eventuales hábitos lingüísticos locales (allí donde fueres, haz lo que vieres… o arriésgate a la extinción), mientras que a los anfitriones/hablantes/oyentes «les convenía» minimizar la carga en los sistemas de memoria y de articulación, manteniendo un repertorio limitado y eficiente de tipos de sonido diferentes. En todo esto no se necesita ningún «esfuerzo consciente», porque las presiones inmediatas son las presiones selectivas de la replicación diferencial.

Un ejemplo convincente de un proceso así lo ofrece un experimento fascinante de Claidière et al. (2014). Un grupo de babuinos, en cautividad pero que pueden desplazarse en un recinto grande, descubren un lugar en que obtienen recompensas si de una matriz de dieciséis cuadrados recuerdan correctamente cuáles cuatro se habían iluminado unos segundos antes. Al principio, ante cuatro cuadrados iluminados situados al azar, cometen muchos errores, pero poco a poco mejoran su actuación y se convierten en buenos «copiadores». A partir de ahí, sus respuestas, incluyendo errores (como mutaciones) se transmiten al siguiente babuino como elementos a recordar, y los patrones a recordar van cambiando gradualmente de las disposiciones iniciales aleatorias a tetrominós conectados más fáciles de recordar, grupos destacados de cuatro cuadrados que forman, líneas, cuadros o formas en L, T o S. Véase la figura 12.1.

FIGURA 12.1. Patrones aleatorios de tetrominós evolucionando********

He aquí la selección darwiniana inconsciente: los babuinos no intentan hacer nada más que obtener una recompensa pulsando los cuatro cuadrados correctos, pero a base de repeticiones, los patrones que se perciben y recuerdan con más facilidad sobreviven, mientras que los demás se extinguen. Los elementos de este ejemplo no tienen ninguna interpretación semántica; son memes bien diseñados (por la replicación diferencial) para propagarse a pesar de no proporcionar ninguna ventaja aparte de una recompensa por copiarse. Tal como afirman los autores, su estudio muestra que

la transmisión cultural en primates no humanos puede dar lugar a la aparición espontánea de comportamientos eficientes, estructurados y específicos de una estirpe, lo que demuestra que compartimos con nuestros parientes más cercanos muchos de los requisitos básicos para crear cultura pseudohumana. (p. 8)

En los primeros días del lenguaje humano, un proceso similar podría aportar unos buenos ladrillos de sonido, memes que eran formas de articular y que luego podían adoptarse fácilmente como ladrillos semánticos: memes que eran formas de comunicar o, en resumen, que eran palabras con significado. Una sobreabundancia de sonidos productivamente generados buscando trabajo (la mejor manera de evitar la extinción) es un «taller de inventos» más productivo que un aluvión de distinciones sin sonidos para expresarlas. Se necesita un cierto tipo de diseñador inteligente para acuñar un neologismo útil y adecuado; actualmente hay muchos ponderados lexicógrafos para este trabajo, pero en las primeras épocas del lenguaje, los sonidos que ya había en circulación pueden haberse adoptado de forma más o menos inconsciente para usarse en ocasiones determinadas, uniendo al instante la experiencia coincidente de un sonido familiar y un elemento destacado (dos ofrecimientos) para formar una nueva palabra, cuyo significado resultaba evidente en el contexto.

De esta manera se puede poblar el léxico con fonología y semántica, pero ¿de dónde viene la gramática? Las articulaciones aisladas y fijadas convencionalmente, como las llamadas de aviso, tienen una variedad semántica limitada: hola, ay, ajá… ¿Cuándo y por qué surgió la distinción entre sustantivo y verbo a partir de las señales o llamadas protolingüísticas, que no diferenciaban expresiones enunciativas de imperativas? Según Hurford, toda lengua tiene, y necesita, una distinción entre tema y comentario (es decir, de qué estás hablando, por un lado, y qué estás diciendo sobre eso, por otro). Sin embargo, algunos idiomas tienen una distinción apenas identificable entre sustantivos y verbos; además, algunos que sí la tienen, siguen la estructura SOV (sujeto, objeto y verbo) mientras que otros siguen la SVO (sujeto, verbo y objeto, como el español: Tomás bebe agua y no Tomás agua bebe) y otros, la VSO (verbo, sujeto y objeto, como el galés); algunos idiomas utilizan mucho las oraciones subordinadas, mientras que otros no. Hurford distingue

entre lo que es obligatorio (y dice por qué; para qué) y lo que es optativo (y puede tener solamente una explicación histórica, un cómo es que) y ordena los elementos de una lengua según su momento de aparición. Las palabras funcionales a menudo descienden de verbos o sustantivos, como las preposiciones durante o mediante o los artículos indefinidos, que acostumbran a ser el número uno (un en español, un en francés, ein en alemán); las palabras de contenido, en cambio, casi nunca descienden de palabras funcionales. Estas pistas históricas, y muchas otras, obtenidas a partir de siglos de erudición fundamentan las hipótesis de Hurford sobre el crecimiento gradual de las gramáticas.

Una de las más reveladoras de estas hipótesis se refiere a la variación en complejidad de los idiomas del mundo (tanto en gramática como en pronunciación). En los pequeños grupos de cazadores-recolectores en que nuestros ancestros vivieron hasta no hace tanto,

la identidad grupal era una fuerza de cohesión social, en la competición con otros grupos […] Era muy probable que cualquier caso de exogamia [apareamiento fuera del grupo] se diera con vecinos cercanos que hablaban un idioma similar. Los hijos se criaban con escaso contacto con forasteros y no había ningún motivo para comunicarse con desconocidos (o muy poco). De este modo, las lenguas de estos pequeños grupos podían evolucionar libremente con sus propias particularidades, sin influencia de otros idiomas. Hay una consistente correlación estadística negativa entre la complejidad morfológica de un idioma y el número de sus hablantes. (Hurford, p. 147)

Dicho de otro modo, una comunidad aislada y pequeña no difiere mucho de una isla, en que la evolución puede generar una gran cantidad de novedades que no podrían sobrevivir en un mundo más grande y competitivo. Los memes que se fijan pueden proporcionar ventajas importantes a sus anfitriones o no; pueden ser simplemente comensales o parásitos que logran prosperar en este refugio, hasta que son forzados a competir con memes invasores.

«El contacto entre adultos que hablan lenguas diferentes tiende a producir variedades de lenguaje en que se elimina la complejidad

morfológica»

(p. 148). La forma evidente de ver la «motivación» de esto es que los hablantes «se ven atraídos» inconscientemente hacia unas expresiones más simples, como respuesta a la falta de comprensión de sus interlocutores. En algunos casos, esta «atracción» puede ser más bien una erosión, una respuesta económica simple a las exigencias físicas de la expresión, un efecto de indolencia o ahorro, que aporta un atajo que otros copian. Pero también se puede lograr no mediante esta caída por la pendiente del ahorro, sino gracias a la mutación y la selección de maneras de hacerse entender, dirigidas por sutiles pistas de expresión facial y otras reacciones de los interlocutores, con una orientación escasa o incluso inexistente por parte de los hablantes. Esta «atracción» gradual a los cambios puede implicar tanto intensificaciones o amplificaciones como simplificaciones para ahorrar energía y tiempo. Como siempre, podemos ver el puente de Darwin que lleva de la «selección inconsciente» a la «selección metódica» y, finalmente, al «diseño inteligente», el cual también se puede considerar que surge por etapas, empezando con pruebas y errores poco imaginativos (como los gritos repetitivos de los ingenuos turistas intentando que los locales entiendan algo en «inglés sencillo»), pasando por innovaciones perspicaces acompañadas de gestos y expresiones faciales hasta llegar a «convenciones» ocasionalistas cuyo significado se entiende fácilmente en el contexto (sacudes la cabeza y sueltas «pequeño» al pescador que no sabe ni una palabra de español y, a partir de ese momento, los peces más pequeños en venta serán pakenio para vosotros dos).

Sin duda, las primeras palabras se asignaron a «cosas para las que teníamos conceptos», lo que significa que estábamos bien equipados por nuestras dotes genéticas innatas o por la experiencia que habíamos tenido hasta el momento para distinguir estos ofrecimientos, tenerlos en cuenta, seguirlos y luego ocuparnos de ellos adecuadamente en las circunstancias normales. Lo que consigue anclar estas competencias en nuestros cerebros aún es un misterio y puede que haya diversas formas de encarnar (¿«encerebrar»?) el dominio de un ofrecimiento. Hurford considera que podemos avanzar aun sin una buena teoría neurocientífica de este hecho:

La relación entre palabras y cosas, es decir, el significado, es indirecta, mediada por conceptos en la cabeza de los usuarios de la lengua. Así, tenemos tres clases de entidades: entidades lingüísticas, como las palabras y las frases, entidades mentales, como los conceptos, y objetos y relaciones mundanas, como perros y nubes, como comer y ser más alto que… En términos del cerebro, sabemos que los conceptos se almacenan ahí, de algún modo, pero no tenemos muy claro cómo […] Si esto os hace sentir incómodos, pensad en la búsqueda de las fuentes del Nilo en el siglo XIX. La gente sabía que el Nilo tenía que tener unas fuentes, como todos los ríos, y sabía que tenían que estar por África central. Finalmente se hallaron. El sintagma las fuentes del Nilo no era una frase sin sentido simplemente porque nadie hubiera localizado aún su referente. (p. 60)[10]

Cuando, entre todos estos ofrecimientos, disponemos de una buena cantidad de maneras de expresar, etiquetar, señalar y decir, surge la oportunidad de algo parecido a una simbiosis: dos ofrecimientos se unen para formar algo nuevo, un concepto en el sentido específicamente humano de una palabra con un significado comprendido. ¿Cómo obtenemos estos diferentes tipos de «cosas», palabras y objetos, en nuestras imágenes manifiestas para que se combinen? Cuando somos niños, descubrimos todas estas palabras y todos estos objetos. Los niños dedican mucho tiempo de sus esfuerzos de exploración a unir cosas (bloques de construcción, muñecos, palos, trocitos de comida, trocitos de basura y cualquier otro objeto manipulable que puedan agarrar).

¿Qué es esto? Puedo mirarlo por todos sus lados, puedo metérmelo en la boca, pegarlo en mi nariz, darle un manotazo, golpear otras cosas con él, apretarlo, dejarlo caer, lanzarlo, agarrarlo con fuerza, ponérmelo en la cabeza… Mientras tanto, puedo balbucear y murmurar y familiarizar mi lengua y mis oídos con estas «cosas» pronunciables. Pronto me podré preguntar cómo se llaman estas cosas y qué significan estos sonidos.[11]

A partir de un caótico batiburrillo de oportunidades, pueden surgir regularidades, solo con una atención puntual y con apenas intención.

Cuando las cosas se vuelven lo bastante familiares, me las puedo apropiar: mi bloque, mi muñeco, mi comida, mis palabras… al inicio no pensadas conscientemente como mías, sino tratadas como posesiones. Con la distinción y el reconocimiento llega la posibilidad de reflexión: valoraciones de que estas dos cosas son la misma y de que esas dos son diferentes pueden dar lugar al reconocimiento del patrón, de orden superior, de igualdad y de diferencia, que, así, pueden convertirse en dos «cosas» adicionales en la imagen manifiesta del niño. Como los ciclos prebióticos que proporcionaron los procesos iterativos a partir de los cuales surgió la vida y la evolución, estas manipulaciones repetidas proporcionan un motor de recombinación a partir del cual se puede construir la densamente poblada imagen manifiesta de una cría humana en proceso de maduración. En el capítulo 14 nos centraremos con más detalle en cómo la imagen manifiesta se vuelve manifiesta para el niño; es decir, en cómo se vuelve parte de su experiencia consciente.

Los cerebros están bien diseñados para captar ofrecimientos de todo tipo y perfeccionar las habilidades para responder adecuadamente ante ellos. Cuando un cerebro se empieza a poblar con memes pronunciables, estos se presentan como oportunidades de control y las capacidades de descubrimiento de patrones del cerebro empiezan a trabajar buscando relaciones entre ellos y los otros ofrecimientos disponibles. Tal como señalamos en el capítulo 9, los niños suelen aprender una media de siete palabras por día durante los primeros seis años de vida, la mayoría de las cuales no se les enseña de forma deliberada señalando («Mira, Carlos, un tenedor. Mira, Ana, gaviotas») y, mucho menos, por definición («Un descapotable es un coche con un techo que se puede plegar»). Los niños adquieren los significados de la mayoría de estas palabras gradualmente, en general sin darse cuenta del inicio de la comprensión, mediante un proceso que se parece muy poco a la formulación y comprobación intencionada de hipótesis, excepto por su capacidad de asegurar resultados con fiabilidad: un análisis estadístico involuntario de los multiformes estímulos con los que se encuentran.[12]

¿Pueden adquirirse las «reglas» gramaticales y morfológicas mediante procesos de abajo arriba, competentes pero sin comprensión? Sí, puesto que nadie aprende la gramática de su lengua materna de arriba abajo, aprendiendo generalizaciones explícitas («En alemán hay tres géneros: masculino, femenino y neutro») o reglas explícitas («En francés los sustantivos concuerdan con los adjetivos en género y número»).

Jackendoff (1994) dramatiza muy gráficamente esta cuestión con su «paradoja de la adquisición del lenguaje»:

Toda una comunidad de profesionales muy cualificados [lingüistas teóricos], aprovechando años de reflexión consciente y compartiendo información, no ha sido capaz de reproducir la hazaña que cualquier niño normal logra a la edad de unos diez años, inconscientemente y sin ayuda alguna. (p. 26)

Sin embargo, hay dos maneras diferentes en que un proceso de abajo arriba puede adquirir esta experiencia: o bien con algún tipo de aprendizaje profundo, un proceso inconsciente de búsqueda de patrones, o bien gracias a la herencia genética. En realidad, es mejor decir que hay dos extremos y la controversia se plantea al determinar en qué lugar del espectro entre estos dos extremos se halla la verdad. En uno de los extremos, la tarea se realiza mediante una competencia de búsqueda de patrones que es completamente general y que no dispone de nada especial acerca del lenguaje; en el otro extremo, tenemos un sistema innato casi por completo (una gramática universal), para el cual solo hay que establecer los «parámetros» de un idioma u otro mediante la experiencia (algo así como personalizar el procesador de textos para adaptarse a tu propio estilo de tecleo, excepto por el hecho de que el niño que está aprendiendo la lengua lo haría inconscientemente). En el ardor de la batalla puede resultar irresistible hacer que parezca que el oponente ocupa uno u otro de los extremos, pero, de hecho, son posibles posiciones intermedias, y más defendibles. El extremo del espectro de «básicamente aprendizaje» ha sido ocupado recientemente por los creadores de modelos y lingüistas que se ocupan del aprendizaje automático (machine learning), sorprendidos por la «gradualidad ubicua» (como hemos visto a menudo en estas páginas), que es tanto un rasgo de las categorías gramaticales como de las especies y subespecies. Pensemos en las diferencias entre:

modismos e idiotismos como a troche y moche o erre que erre, que son bastante impenetrables a un análisis interno;

tener malas pulgas o estirar la pata, cuyo significado no se puede derivar fácilmente analizando sus constituyentes (y, por tanto, también tienen que aprenderse como elementos léxicos autónomos);

cantar las cuarenta o tener un as en la manga, que se pueden analizar si sabes algo sobre juegos de cartas, pero son lo bastante familiares para aprenderse y usarse sin tal conocimiento;

personaje protagonista, sentimientos contradictorios o sin remedio, que son «elementos prefabricados» (Bybee 2006), «convencionalizados pero [con significados] predecibles de otros modos», y

dónde se halla la verdad y proceso de abajo arriba, que cualquiera que conozca el significado de los componentes, puede comprenderlos en su contexto.

Estas gradaciones llevan su historia por bandera (un bonito caso a medio camino), ya que a menudo podemos reconstruir el proceso de gramaticalización que toma combinaciones replicadas con frecuencia y las solidifica gradualmente en unidades que pueden replicarse por sí solas como unidades combinatorias. Pero incluso si «en principio» todas las regularidades gramaticales pudieran evolucionar en la comunidad lingüística e instalarse en los individuos sin ayuda de los genes (el extremo de «básicamente aprendizaje»), se puede plantear un argumento de peso para una cierta contribución innata.

El argumento más influyente para un Dispositivo de Adquisición del Lenguaje, o DAL (Chomsky 1965), es el que se basa en la «pobreza de los estímulos», que afirma que un niño humano no oye el suficiente lenguaje gramatical (y lenguaje agramatical corregido: prueba y error) durante los primeros años de vida para obtener los datos necesarios que le permitan construir una «teoría» de su gramática. Los pájaros que nunca han visto construir un nido, pueden construir uno rápidamente, practicable y específico de su especie, gracias a su competencia innata de construcción de nidos; los niños a los que nunca se les han enseñado las sutilezas de los adverbios y las preposiciones pueden estar pertrechados para el lenguaje de la misma manera. Como la competencia gramatical tiene que provenir de algún sitio, prosigue el argumento, tiene que ser innata por lo menos en parte, un conjunto de principios, reglas o

restricciones que permite (de hecho, que obliga) a los niños a reducir su búsqueda en el Vasto espacio de posibilidades en un proceso de prueba y error dirigido y no aleatorio. Una grúa así (véase el capítulo 4) explicaría la fascinante facilidad con que los niños van a parar a la gramática de su lengua materna: están escogiendo de un conjunto comparativamente pequeño de lenguajes aprehensibles, restringido gracias a las restricciones integradas del DAL. Pero Chomsky hizo que, en algunos ámbitos, se lanzara un velo de sospecha sobre esta afirmación, a causa de su terca resistencia ante cualquier intento de explicación del DAL mediante la selección natural. El punto de vista de Chomsky hacía que el DAL pareciera más un gancho celeste que una grúa, un salto inexplicable que se produjo milagrosamente en el espacio de diseño, un regalo de Dios, no el resultado de una trabajosa I+D por parte de la evolución durante muchas generaciones.

Quizá como respuesta, en parte, a estas críticas, Chomsky (1995, 2000, 2000b) revisó radicalmente su teoría, y desde entonces ha defendido el programa minimalista, que descarta todos los mecanismos innatos y restricciones de su versión anterior y propone que todo el trabajo que se supone que tenían que hacer podía ser efectuado por un único operador lógico, denominado ensamble (en inglés, merge) (véase Hauser, Chomsky y Fitch 2002 para una célebre defensa de esta postura en la revista Science, y Pinker y Jackendoff 2005 para una refutación detallada con referencias útiles a la riada de publicaciones críticas provocadas por el trabajo). El ensamble es un operador combinatorio multipropósito, que «fusiona recursivamente dos elementos (palabras o frases) en un árbol binario que lleva la etiqueta de uno de ellos» (Pinker y Jackendoff 2005). Así, según Chomsky, el único talento cognitivo necesario para el lenguaje, la competencia que poseen los seres humanos pero no el resto de animales, es la operación lógica conocida como recursión, que puede instalarse sin la extensa I+D de la selección natural, en una especie de accidente cósmico de una sola etapa, un objeto que hemos hallado (¡afortunados de nosotros!), no un recurso que ha evolucionado.

En matemáticas e informática, las funciones recursivas son funciones que pueden tener como argumento a ellas mismas, lo que significa que tras haber aplicado la función una vez a una determinada entidad y, así, haber obtenido una nueva entidad, se puede aplicar la misma función a esta nueva entidad acabada de obtener, y así sucesivamente, ad infinitum. Este giro permite generar múltiples anidamientos, como muñecas rusas.

La verdadera recursión es un poderoso concepto matemático, pero hay imitadores baratos que pueden hacer casi el mismo trabajo sin ser estrictamente recursivos. El ejemplo habitual de recursión en el lenguaje natural es la incrustación o encaje de oraciones subordinadas: «He aquí al gato que mató al ratón que se comió el queso que estaba en la alacena que decoraba la casa que construyó Pedro». Está claro que, en principio, se puede construir una frase inacabable añadiendo frases hasta el infinito, y uno de los sellos distintivos del lenguaje es su, en teoría, infinitud. No hay ninguna frase gramatical en español que sea la más larga posible. Ahora bien, hay otras formas de llegar a este infinito: por ejemplo, «Tomás se comió una alverja y luego otra alverja y luego otra alverja y luego otra alverja y luego…» también genera una frase infinitamente larga (e infinitamente aburrida), pero esto no es más que iteración, no recursión. E incluso cuando hay recursión implicada en el lenguaje natural, hay una frontera clara que limita cuántas recursiones pueden controlar unos hablantes competentes. Si puedes hacer siete recursiones, eres un verdadero cinturón negro del análisis sintáctico, pero me pregunto si te das cuenta de que intento persuadirte con mi informal demostración de que incluso tú mismo tienes ciertos límites que supongo que debe estar alrededor de los siete, aunque esta frase tiene solo seis. Además, parece ser que por lo menos existe un idioma, el pirahã en la Amazonía brasileña que no tiene subordinaciones en absoluto (Everett 2004), y también hay funciones comparativamente simples que parecen recursivas hasta que intentas aplicarlas a ellas mismas más de una vez (o de dos… o de n veces, para n finito). Microsoft Word, por ejemplo, dispone de las operaciones tipográficas de superíndice y subíndice, tal como se muestra:

basepotencia

y

hembrahumana

Pero intentemos añadir otro superíndice a basepotencia: debería funcionar, ¡pero no es así! En matemáticas, puedes elevar potencias elevadas a potencias elevadas a potencias hasta el infinito, pero no

puedes hacer que Microsoft Word las represente (otros sistemas de composición de textos sí pueden hacerlo, como LaTeX). Después de todo esto, ¿estamos seguros de que los lenguajes humanos utilizan una verdadera recursión o quizá algunos, o todos, son más bien como Microsoft Word? ¿Podría ser que nuestra interpretación de las gramáticas como operaciones recursivas no fuera más que una elegante idealización matemática de las «partes móviles» reales de una gramática?

Pinker y Jackendoff presentan un argumento de peso, con gran cantidad de pruebas empíricas, contra las ideas que Chomsky ha esgrimido para respaldar su programa minimalista y también contra las ideas que afirman que de él se deducen.[13] Muestran que la afirmación de que el ensamble puede realizar todas las tareas efectuadas por los sistemas anteriores o es falsa o es estúpida, ya que luego los chomskianos reintroducen, como construcciones aplicadas mediante el ensamble, la mayoría de características de diseño que el minimalismo había expulsado. Así, irónicamente, si ignoramos el antagonismo de Chomsky a la hipótesis de que la selección natural es responsable del DAL, su propuesta de ensamble elemental parece un candidato plausible para una primera adaptación de la gramática, a partir de la cual habrían descendido todos los memes gramaticales posteriores.

Es más, también podemos conjeturar, con ciertos visos de plausibilidad, que el ensamble no fue un paso de gigante accidental, no fue un salto en el diseño, sino un desarrollo gradual a partir de versiones más concretas del ensamble, como las que se pueden ver en las manipulaciones de niños (y adultos) actualmente: poner un bloque encima de un bloque encima de un bloque; usar un martillo grande para hacer un martillo más pequeño para hacer un martillo más pequeño; poner bayas en un montón, poner el montón de bayas en un montón más grande, poner el montón más grande en un montón aún más grande; colocar un montón en una taza en una caja en una bolsa, etc. Pero ¿son estos procesos una recursión verdadera? La pregunta está mal planteada, como la pregunta ¿son los homininos implicados verdaderos Homo sapiens? Sabemos que en la evolución las transiciones graduales son la norma, y una aparición gradual de (algo parecido a) la recursión real, o por lo menos lo bastante real para el lenguaje natural, sería un buen trampolín, siempre que pudiéramos identificarlo.

A lo largo de los años, muchos de los argumentos en lingüística teórica, en su vertiente más abstracta, han implicado estrictas «condiciones necesarias y suficientes» o «criterios» o «marcadores de diferencia» (entre verbo y sustantivo, entre tema y comentario, entre frase y oración y, en particular, entre la lengua A y la lengua B). En resumen, los lingüistas a menudo se han sentido tentados de proponer esencias. Pero ¿cuándo dos hablantes hablan exactamente la misma lengua (si es que lo hacen alguna vez)? Podemos decir que cada hablante real tiene un idiolecto, un dialecto con un único hablante nativo, y si bien tu idiolecto del español y el mío puede que sean casi indistinguibles (lo que explica nuestra capacidad para comunicarnos tan eficientemente), cuando no nos ponemos de acuerdo en si una determinada frase es gramatical o no, o si una clase particular es un subconjunto de otra clase, no hay ninguna autoridad final a la cual recurrir, aparte de la ley de la mayoría. Y ¿cuándo debemos considerar que una mayoría local desautoriza a una mayoría global? Un lingüista y filósofo que conozco, en una ocasión insistía con firmeza en que ni Joseph Conrad ni Vladimir Nabokov hablaban inglés, estrictamente hablando; ¡solo los hablantes nativos de inglés hablan inglés! Pero ¿cuáles de ellos? ¿Los de Londres, los de Brooklyn, las hablantes del valleyspeak de Los Ángeles, los neozelandeses? Los dilemas a los que se enfrenta un taxonomista lingüístico son sorprendentemente parecidos a los que asolaban a cualquier experto taxonomista predarwiniano al definir géneros, especies y variedades. Si seguimos a los darwinianos adoptando el pensamiento poblacional, estos problemas se muestran como artefactos de un esencialismo inapropiado. Ahora bien, poblaciones ¿de qué? De memes, por supuesto.

En 1975, Chomsky defendió el DAL destacando que «un niño normal adquiere este conocimiento [de la gramática] […] sin ninguna formación específica. Luego puede utilizar sin esfuerzo una compleja estructura de reglas específicas y principios rectores para expresar sus pensamientos». Si reconsideramos el ensamble de Chomsky, o algo parecido, como un candidato temprano a una innovación transicional hacia los lenguajes modernos, podremos reconciliar al primer Chomsky y al Chomsky posterior diciendo que la «compleja estructura de reglas específicas y principios rectores» no son tanto reglas explícitas como patrones profundamente incorporados de maneras de hablar, consistentes en una serie de mejoras forjadas por la evolución, tanto cultural como genética, como respuesta al éxito de los protolenguajes. Tal como hemos visto una

y otra vez a lo largo de estos capítulos, nosotros, como otros animales, somos los beneficiarios inconscientes de sistemas diseñados con brillantez para lograr objetivos que compensan la I+D invertida, lo que representa otro caso de competencia que ha evolucionado sin requerir apenas comprensión.

El origen evolutivo del lenguaje es un problema no resuelto, pero no irresoluble; diversos trabajos experimentales y teóricos han avanzado formulando hipótesis comprobables acerca de procesos evolutivos graduales e incrementales, tanto culturales como genéticos, que podrían transformar los talentos más primitivos de nuestros ancestros en la habilidad y prolijidad verbal de los modernos usuarios de la lengua. La llegada del lenguaje preparó el terreno para otro gran momento en la historia evolutiva: el origen de la comprensión.

En el siguiente capítulo veremos que, a medida que aumentaba la competencia lingüística, no solo se aceleró la evolución cultural, sino que el propio proceso (de evolución cultural) pudo evolucionar hacia algo menos darwiniano, menos de abajo arriba, lo que allanó el camino a la comprensión de arriba abajo, uno de los frutos más recientes del árbol de la vida, y al inicio de la era del diseño inteligente. La creatividad de los seres humanos individuales se puede contemplar como un reflejo, a alta velocidad y concentrado, de los procesos de I+D que la crearon.

******** Las representaciones pushmi-pullyiu son representaciones con contenido enunciativo e imperativo a la vez. Su nombre procede de un personaje de la serie de cuentos infantiles Doctor Doolittle, muy populares en los países de lengua inglesa. (N. del t.)

******** Se puede visualizar la imagen de la figura utilizando un lector de códigos QR o bien accediendo a http://pasadopresente.com/images/PDF-LIBROS/delasbacteriasabach_imagenesqr.pdf

13

LA EVOLUCIÓN DE LA EVOLUCIÓN CULTURAL

No puedes hacer grandes trabajos de carpintería solo con tus manos, como tampoco puedes hacer grandes tareas de pensamiento solo con tu cerebro.

B

O DAHLBOM

UNOS INICIOS DARWINIANOS

Recordemos nuestro avance del capítulo 7. Generaciones de naturalistas han determinado que los animales progenitores pueden impartir habilidades y preferencias a su descendencia, sin el beneficio de las instrucciones verbales; estas «tradiciones animales» (Avital y Jablonka 2000) son una especie de evolución memética, pero los memes de los animales no suelen facilitar oportunidades para obtener más memes, tal como sí lo hacen las palabras. No se produce el efecto de bola de nieve acumulativo que el lenguaje posibilita y, tal como subrayamos en el capítulo 12, hay hechos relevantes desde el punto de vista ecológico, e incluso mortales en potencia, sobre circunstancias perceptualmente ausentes que no se pueden expresar sin el lenguaje (como por ejemplo, qué hacer si te encuentras con un oso). La referenciación remota es, tal como arguyen Bickerton y otros, un paso gigantesco en el espacio de diseño.

Ha llegado el momento de sustanciar la afirmación central de la parte II de este libro:

La cultura humana se inició como un fenómeno profundamente darwiniano, con competencias sin comprensión que generaban diversas estructuras valiosas de modo parecido a como las termitas construyen sus castillos. Luego, durante los siguientes cientos de miles de años, la exploración cultural del espacio de diseño se desdarwinizó poco a poco, creando grúas que se podían emplear para construir nuevas grúas que permitían alzar todavía más grúas, en un proceso formado por una comprensión cada vez más potente.

¿Hasta dónde pudo llegar Homo sapiens con una comprensión simiesca? Algún día podremos saberlo si conseguimos pruebas suficientes para determinar unas fechas aproximadas para hitos tan relevantes como la caza y el carroñeo cooperativos, el dominio del fuego, la construcción de viviendas, la elaboración de herramientas y armas, pero sabemos que todo cambió cuando apareció el lenguaje. En el diagrama (figura 13.1), la cultura humana empezó en la esquina inferior izquierda y, gradualmente, se expandió para incorporar bolsas de comprensión creciente (subiendo por el eje vertical y), más control de arriba abajo (creciendo por el eje horizontal x) y una búsqueda dirigida más eficiente (a lo largo del eje z). Yo planteo que la cultura humana fue avanzando en estas tres dimensiones en dirección diagonal, desde el darwinismo puro hasta la cima (inalcanzable) del diseño inteligente, porque los fenómenos dispuestos a lo largo de cada dimensión están ordenados por su uso cada vez más intensivo de información semántica y, por tanto, de mejoras de diseño, que exigen más I+D, más recogida, aprovechamiento y perfeccionamiento, basándose, como siempre, en los fundamentos dispuestos por la I+D previa (más darwiniana). Algunos pasos son más largos que otros, pero no necesitamos lo que Freeman Dyson denominó un regalo de Dios para llegar donde nos hallamos hoy en día. Nos fijaremos con detalle en las controversias acerca del estado actual de la evolución cultural después de repasar los fenómenos darwinianos, semidarwinianos y hemisemidarwinianos que nos han llevado a ella.

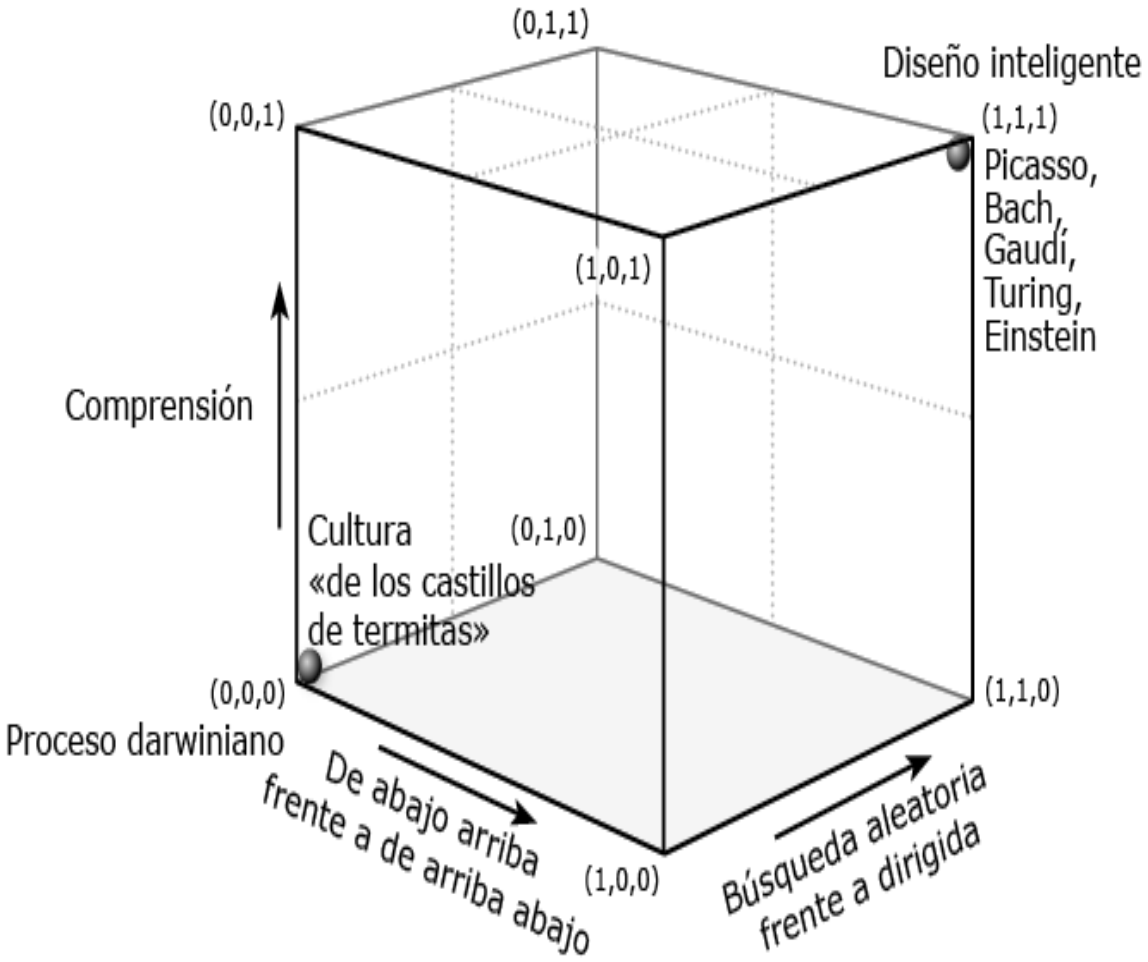


FIGURA 13.1. Espacio darwiniano

En el capítulo 12 hemos preparado el camino: los memes, como los virus, son simbiontes que dependen de la maquinaria reproductora de sus anfitriones, a la que explotan para sus propios fines. Así, para que haya una explosión demográfica de memes, tiene que haber habido un instinto preexistente de imitar o copiar (o en evolución conjunta), que «se costearía» al proporcionar ciertas ventajas de aptitud (genética) a nuestros ancestros. Parece ser que los antecesores de nuestros parientes no humanos más cercanos no disfrutaron de las condiciones propicias para ello, o no durante el tiempo suficiente, y así no pudieron unirse a nosotros en este trampolín. Por ejemplo, los chimpancés y los bonobos no muestran el interés, la atención centrada ni el talento imitativo necesarios para prender la hoguera cultural acumulativa que nos diferencia de los

demás homínidos.[1] Investigaciones futuras pueden decantar la balanza a favor de una u otra de las hipótesis rivales acerca de cómo se produjo esto, pero creo que ya hemos visto lo suficiente como para quedar convencidos de que no se necesitó ningún milagro para que nos pusiéramos en camino.

Una vez nos pusimos en marcha, nos convertimos en simios con cerebros infectados por memes. Como los virus y otros simbiontes que se han establecido a billones en nuestros cuerpos, los invasores evolucionaron para ser unos reproductores más eficaces, manteniéndose frente a los competidores dentro de los cuerpos, venciendo en las competiciones de propagación y extendiéndose a nuevos anfitriones. Entre los parásitos, debía haber los suficientes mutualistas y comensales para no matar a sus anfitriones, aunque es perfectamente posible que oleadas de infecciones por memes hayan hecho exactamente eso antes de que una oleada por fin resultara lo bastante benigna para asegurar una buena posición a largo plazo. Podemos imaginar a un grupo de potenciales ancestros aquejados de una obsesión por la danza tan abrumadora que les hizo abandonar las tareas de caza y recolección y los llevó a una situación de inanición; o también a un desafortunado linaje cuyo paso a la edad adulta implicara una circuncisión ritual que cortara cada vez más, en una versión fracasada de la cola del pavo real. De modo parecido, en los primeros días de la vida, esta pudo haber aparecido varias veces, solo para apagarse una y otra vez, hasta que uno de estos orígenes consiguió suficientes características adecuadas (lo bastante adecuadas) para sobrevivir indefinidamente.[2]

Los primeros memes, tanto si eran protopalabras pronunciables como si eran mudos hábitos conductuales de otro tipo, eran sinántropos, aún sin domesticar, y tenían que ser especialmente «infecciosos» hasta que estuviera a punto el conjunto de genes y memes que permitiera mejorar las rutinas de copia. Es probable que solo una minoría de ellos tuviera alguna capacidad funcional aparte de su poder reproductor. Los malos hábitos, pero pegadizos, serían un precio que valía la pena pagar a cambio de los hábitos realmente buenos y una vez estos buenos hábitos estuvieran a punto, habría tiempo suficiente para que la I+D genética y cultural pudiera mirar hacia atrás y hacer un poco de limpieza de los excesos, de forma parecida a cómo nuestra micromaquinaria genética ha creado medidas para ocuparse de los alborotadores del genoma, como parásitos genómicos, transposones y distorsionadores de la segregación (véase el capítulo 12). En cuanto la comunicación verbal se convirtió no solo en un

Buen Truco, sino en un talento obligatorio para nuestra especie, habría una presión selectiva constante a favor de modificaciones orgánicas que mejoraran o racionalizaran el proceso de adquisición del lenguaje. Entre estas importantes innovaciones destaca la de la altricialidad (la infancia prolongada), que amplió el tiempo durante el cual la descendencia depende de los progenitores para su protección, alimentación y (en absoluto involuntariamente) educación. Este gran aumento en la interacción «cara a cara» fue potenciado aún más gracias al control de la mirada, lo que, a su vez, permitió la atención conjunta y, tal como señala Tomasello, la intención conjunta (el único otro mamífero que se implica regularmente en el control de la mirada es el perro doméstico, y vigila la mirada de su dueño humano, no la de otros perros; véase, por ejemplo, Emery 2000 y Kaminski 2009). Podemos comparar la esclerótica blanca de los humanos con las oscuras escleróticas que rodean las pupilas de los otros simios; es posible que estemos ante una adaptación reciente para mejorar el control de la mirada, un bello ejemplo de coevolución cultural/genética, una respuesta genética a un nuevo comportamiento diseñado para potenciar la transmisión de los memes (Frischen et al. 2007 presentan una extensa descripción).

¿A qué se parecía todo esto para estos pioneros anfitriones de los memes? ¿Era ser como algo? Sus cabezas estaban repletas de herramientas, recursos, juguetes y basura, todos «hechos de información» y ocupando tejido cerebral y gastando energía del cerebro; pero ¿cuáles de entre estos invasores se daban a conocer a sus anfitriones? Los gérmenes y virus que ocupan nuestros cuerpos pasan desapercibidos excepto cuando provocan cambios somáticos a gran escala: dolor, palpitaciones, vómitos, estornudos, mareo, etc. Tal como vimos en el capítulo 4, la ontología implícita en el diseño de un ascensor explica su competencia para controlar sus actividades sin que el ascensor tenga que comprender qué está haciendo, sin que tenga que ser consciente en absoluto de su ontología, excepto en su sentido más restringido (solo tiene que ser sensible a ella, distinguir sus elementos y reaccionar de formas diferentes a situaciones diferentes). Esta es la ontología que se les ha ocurrido a los ingenieros para el ascensor que han diseñado. Las ontologías de nuestros ancestros incluían palabras y otros memes: podían usar estas palabras de forma competente y se beneficiaban de tener palabras sin darse cuenta todavía de que estaban usando palabras, sin que las palabras les fueran manifiestas como palabras en su imagen manifiesta. En un sentido restringido, se percataban de las palabras (eran sensibles perceptivamente a ellas y respondían de formas diferentes a diferentes palabras, como el

ascensor) pero no tenían por qué percatarse de que se estaban percatando de ellas. Perceptivamente, las palabras podían pasar de un individuo a otro, penetrando y asentándose en sus cuerpos, como las vitaminas o la flora intestinal: simbiontes valiosos cuyo valor no dependía de que fueran identificados ni apreciados por sus anfitriones.

Aun así, el hogar natural de los memes es nuestra imagen manifiesta, no nuestra imagen científica (donde se hallan las vitaminas y la flora intestinal). Los memes solo pueden viajar por el aire si generan efectos perceptibles por los humanos, de modo que, en general, están disponibles para percatarse de ellos, aunque muchos se introduzcan de forma inadvertida e inesperada.[3] Los memes, a diferencia de los virus y los microbios, son ofrecimientos que estamos capacitados para percibir, reconocer, recordar y responder a ellos adecuadamente desde el principio. Son elementos acreditados en nuestra ontología, aunque habitualmente sean «invisibles» para otros animales, aunque sean elementos que no forman parte de sus ontologías. En este punto debemos recordar la distinción de Sellars entre la imagen manifiesta y la imagen original más primitiva (véase el capítulo 4, nota 4, p. 379). Cuando nos percatamos de nuestros memes y empezamos a poseerlos y a reflexionar sobre ellos, pasamos de la imagen original a la imagen manifiesta, al mundo en que vivimos y del cual sabemos que es el mundo en que vivimos.

LAS MOTIVACIONES FLOTANTES DE LA COMUNICACIÓN HUMANA

La competencia sin comprensión es tan omnipresente en la vida humana como en los animales, las bacterias y los ascensores, pero tenemos tendencia a pasar por alto esta posibilidad y a adjudicar a los actores inteligentes implicados una apreciación de las motivaciones de las acciones humanas completadas con éxito. Esto no es sorprendente. Después de todo, tendemos a atribuir más comprensión a lobos, aves y abejas que la que podemos justificar con pruebas: el uso del enfoque intencional para interpretar el comportamiento en todas las especies conlleva una suposición tácita de racionalidad, y ¿qué racionalidad puede ser si no es la de los individuos que actúan? La propia idea de

motivaciones flotantes es una extraña inversión del razonamiento, tal como podemos ver siguiendo una serie de controversias que se dieron durante años, provocadas por un famoso trabajo en la filosofía del lenguaje.

Uno de los hitos de la filosofía del lenguaje en el siglo XX es la descripción de H. P. Grice (1957, 1968, 1969, 1989) de las condiciones necesarias para la comunicación, o lo que él denominó «significado no natural». La afirmación central de Grice era una definición en tres partes de lo que representa para una persona significar algo haciendo algo. Tal como clarificaron Strawson (1964) y otros, para significar algo haciendo x, S debe tener la intención de

que x produzca una cierta respuesta r en una determinada audiencia K;

que A reconozca la intención (1) de S, y

que el reconocimiento de (1) por parte de A actúe, por lo menos en parte, como la razón de A para su respuesta r.

Este elegante conjunto recursivo tiene muchas implicaciones. A puede aprender mucho de algo que haga S sin que S quiera significar algo al hacerlo. Por ejemplo, 1) si S grita de dolor o se desmaya, A puede recibir la información de que S tiene algún problema, pero puede que S no haya tenido la intención de informar a A, puede ser un comportamiento involuntario de dolor. 2) Si S sí tiene la intención de informar (o informar erróneamente o desinformar) a A pero no tiene la intención de que A se dé cuenta de ello (piénsese en un soldado que finge estar enfermo para ser enviado al hospital de campaña o en el pájaro que simula estar herido), la situación no cuenta como comunicación. Finalmente, 3), si no es el reconocimiento por parte de A de esta intención comunicativa lo que fundamenta la creencia de A, tampoco se trata de un caso de comunicación, por muy informativa que sea. Por ejemplo, mi esposa deja encendida la luz de la cocina para asegurarse de que yo vea los platos sucios que tengo que lavar; quiere que lo capte, pero no está necesariamente enviándome un mensaje; de hecho, me lo enviará en

cuanto yo no reconozca su intención de llamar mi atención sobre mis tareas.

El análisis de Grice parece sincero para la mayoría de los que se han topado con él y ha inspirado una riada de trabajos teóricos posteriores, buena parte de los cuales presentan una retahíla supuestamente inacabable de ingeniosos contraejemplos y problemas (véase Azzouni 2013 para una exposición detallada y perspicaz). Centra la atención en las sutiles diferencias con que los humanos podemos interactuar y, en el ínterin, subraya las sutilezas que parecen estar ausentes en los sistemas de comunicación animales. Pensemos en la anécdota del cercopiteco que emite una falsa alarma de águilas justo cuando su tribu está a punto de ser derrotada en un enfrentamiento con una tribu rival, con lo que provoca un cese de las hostilidades y permite recuperar el terreno perdido: ¿es esto un ejemplo de engaño (intencional) o un caso de pura suerte? Las gacelas saltarinas no necesitan que los leones cazadores diferencien entre un fanfarroneo deliberado y una euforia natural; en ambos casos, los leones las dejarán tranquilas si saltan. La tentadora conclusión que se puede extraer (y muchos, como yo mismo, hemos aceptado esta tentación) es que las condiciones anidadas de Grice muestran que el comportamiento comunicativo humano es de un orden diferente que el mero comportamiento animal, más simple, pues implica una intencionalidad de cuarto orden, por lo menos:

S (1) tiene la intención de que A (2) identifique que S (3) tiene la intención de que A (4) crea que p.

Grice no solo ha inspirado a filósofos; los antropólogos y psicólogos Dan Sperber y Deirdre Wilson (1986) usaron a Grice como punto de partida para una teoría rival del significado, pero no pusieron reparos en apreciar las aportaciones generales de Grice:

Desde un punto de vista psicológico, la descripción de la comunicación en términos de intenciones e inferencias también tiene todo el sentido. La

atribución de intenciones a los otros es un rasgo característico de la cognición y la interacción humanas.

Ruth Millikan (1984), en cambio, al principio tuvo una actitud desdeñosa:

Si creyera que Jack el Destripador está escondido debajo de mi cama, no me metería en la cama y me quedaría dormida al instante. Si me meto en la cama, está claro que no creo la afirmación «Jack está debajo de mi cama»; pero de ello no se deduce que crea «Jack no está debajo de mi cama». De hecho, quizá nunca he oído hablar de Jack. Igualmente, del hecho de que si tuviera alguna razón para creer que un hablante no tenía la intención de que yo acatara una orden entonces es muy probable que no la acatara, no se desprende que en casos normales de acatamiento de órdenes yo creyera que el hablante tiene la intención del acatamiento. (p. 61)

En la página siguiente matiza un poco su postura:

En primer lugar, hay el problema de qué es exactamente tener una determinada creencia o intención. Luego está el problema de qué es usar una creencia para hacer algo, en oposición, por ejemplo, a solo tenerla mientras se hace algo. También debemos preguntarnos si hay necesidad alguna de interpretar las intenciones y creencias griceanas como cosas que realmente se usan mientras se habla y se comprende o si pueden tener otro estatus y sin embargo seguir haciendo el trabajo que los teóricos griceanos quieren que hagan, de distinguir un tipo de «significado» que tienen todos los signos no naturales.

Millikan, que se destaca como una famosa no griceana, concluye, finalmente:

Solo las creencias griceanas que realmente se usan en el proceso de hablar y comprender pueden ser de interés para la teoría del significado. (p. 66)

Esto no ha disuadido a otros de adoptar el programa griceano de un modo u otro. No hace mucho, mi colega Jody Azzouni (2013) se ha enfrentado sin rodeos al análisis griceano, y vale la pena destacar el esfuerzo que hace para ablandar a los lectores ante su inminente asalto:

Mi principal objetivo en este libro no es refutar los enfoques griceanos y neogriceanos de una vez por todas. El griceanismo está demasiado extendido y sus diversos defensores son demasiado diferentes en sus enfoques particulares. Además [añade en una nota al pie], ¿quién quiere hacerse tantos enemigos de golpe? (p. 4)

Para los estudiantes de literatura, el paciente e imaginativo análisis de Azzouni es lectura obligada,[4] pero aquí quiero subrayar solo una de las divertidas bombas de intuición que fabrica para sorprendernos con un punto de vista no griceano:

Imaginemos dos personas en una isla, cada una hablante de un idioma sin relación con el otro. Él sostiene un objeto y empieza a hacer mímica mientras ella intenta descubrir qué está intentando decir. Difícilmente puede una situación ser más griceana. Ella identifica que él tiene una intención comunicativa; él identifica que ella identifica que él tiene una intención comunicativa (y así sucesivamente, quizá). A partir de esto, del contexto y de algún otro conocimiento básico común (hace tiempo que han estado haciendo estas cosas y quizá han elaborado un repertorio de significados), ella intenta determinar qué está intentando comunicar y, en consecuencia, qué significa el objeto que sostiene.

Difícilmente puede ser una situación más griceana, pero, siendo honestos, tampoco puede ser mucho más atípica y frustrante. La comunicación griceana real es bien fastidiosa (como cualquiera que se haya visto obligado a implicarse en ella podrá corroborar con rabia). (p. 348)

El ejemplo de Azzouni muestra que la comunicación cotidiana es muy diferente de la comunicación griceana, pero en tal caso, tenemos un enigma: ¿cómo es que Grice convenció a tantas personas de que había dado con algo original e importante? ¿Acaso provocó una alucinación intelectual colectiva? Podríamos decirlo así, en efecto. La siguiente frase de Azzouni señala la dirección correcta: «El lenguaje ordinario puede haberse originado, en términos evolutivos, a partir de sucesos parecidos a los sucesos de comunicación griceana real, pero desde entonces han cambiado muchas cosas y, especialmente, en nuestros cerebros» (p. 348).

Lo que hizo Grice, sin darse cuenta, es aplicar la ingeniería inversa a la comunicación humana, alegando las motivaciones naturales que eones de evolución genética y cultural descubrirían de forma natural una vez se había establecido el Buen Truco básico de usar palabras como herramientas. Del mismo modo que cuando explicamos el fenómeno del salto de rebote de las gacelas y del polluelo cuco asesino, la forma de hablar que nos viene a la cabeza de manera natural es el enfoque intencional, que trata a los agentes como agentes racionales cuyas razones se manifiestan de forma aparente, sin fijarnos en que estas razones no tienen por qué ser comprendidas, percatadas y ni tan solo representadas por esos agentes. El punto central de Grice era, o debería haber sido, que los comunicadores humanos (en cuanto captan el enorme potencial de los recursos que tienen a su disposición) tienen la competencia para aprovecharse de todas estas características, y la competencia para evitar que otros se aprovechen de ellos aprovechándose de esas mismas características.

Recordad que mi propuesta es que la adquisición de un lenguaje (y, en general, de los memes) es muy parecida a la instalación de una potente aplicación informática prediseñada, como Adobe Photoshop, una herramienta para profesionales con muchas capas que la mayoría de usuarios aficionados nunca van a usar. En la comunicación humana hay mucha variación y la mayoría de usos del sistema son rudimentarios, rutinarios, guiados por hábitos que no son comprendidos por la mayoría de observadores (y autoobservadores). Pero las herramientas permiten usos muy sofisticados. Algunas personas son manipuladores natos, formadores de impresiones, maestros de la falsedad y de las adulaciones subliminales, mientras que otros son directos, ingenuos, imprudentes (podríamos decir que son usuarios novatos de las herramientas), pero

ninguno de estos tipos de personas tiene por qué comprender las razones por las que sus recursos comunicativos cotidianos ofrecen todas las opciones que ofrecen.

Nuestros sistemas de comunicación (verbal) han sido brillantemente diseñados, mediante un proceso de selección cultural natural que los ha dotado de un conjunto de rasgos que los nuevos usuarios adquieren gradualmente y quizá nunca llegan a comprender en profundidad. Pueden desarrollar una aguda sensibilidad a los riesgos, peligros y oportunidades de la comunicación que no son capaces de analizar. Pueden «oler» una mentira o una falsa modestia, por ejemplo, sin tan siquiera ver el rostro del hablante u oír su tono de voz; la elección de las palabras puede «dar mala espina» de algún modo impenetrable a sus esfuerzos de análisis. Otros hablantes de una lengua pueden tener facilidad de palabra u otros dones (la capacidad de consolar, de persuadir, de seducir, de divertir, de inspirar). Todo esto «se da de forma natural» en algunas personas y está más allá de las capacidades del resto. Algunos individuos, especialmente con trastornos del espectro autista «de alto funcionamiento», como Temple Grandin, consiguen concebir, con grandes dosis de esfuerzo e ingenio, una verdadera «teoría de la mente» que les ayuda a interpretar el calidoscópico mundo social que la mayoría de nosotros «percibimos directamente». También estamos adquiriendo poco a poco comprensión teórica, y un genuino entendimiento, de las razones de los rasgos de nuestros sistemas de comunicación. Los que destacan en este campo suelen ofrecer sus servicios como diseñadores inteligentes de comunicación (instructores de oratoria, consultores de mercadotecnia o publicistas), pero tal como podemos confirmar a partir de otros ámbitos (el jazz, sin ir más lejos), la teoría no está a la altura del oído, incluso siendo una buena teoría. Si alguna vez habéis visto a algún brillante científico con síndrome de Asperger elaborando actos de habla y otros trucos sociales con una velocidad encomiable sin poder ocultar los signos reveladores del ardid, podréis apreciar que, si el análisis de Grice fuera una teoría de ejecución,[5] sería aplicable solo a una minoría de hablantes.

Los proyectos de aplicar la ingeniería inversa a la actividad humana a través del enfoque intencional han sido concebidos a menudo por filósofos, aunque no con tal apariencia. G. E. M. Anscombe, en su oscura obra maestra Intención (1957), hablando sobre la descripción de Aristóteles del silogismo práctico, afirma:

Si se supone que la explicación de Aristóteles tiene que describir los procesos mentales reales, en general sería bastante absurda. El interés de la explicación radica en que describe un orden que está presente siempre que se realizan acciones con intenciones.

¿«Un orden que está presente»? ¿Dónde? No en el cerebro, no en el entorno, no representado en ningún sitio. Me permito sugerir que estaba reconociendo a Aristóteles el mérito de descubrir las motivaciones flotantes que «gobiernan» nuestro comportamiento práctico más racional sin que estén siempre representadas.

Así, incluso cuando las palabras han sido incautadas como herramientas domesticadas y, en consecuencia, han pasado a formar parte de la imagen manifiesta, es posible que no nos figuremos la motivación flotante del diseño de algunas de nuestras prácticas. Podemos considerar que Grice lo resolvió y vio el «orden que está presente» cuando las personas se embarcan en significados no naturales y simplemente lo presentó como una descripción de sus estados intencionales del momento. Así de natural es nuestra práctica de dotar a las personas de excesivas razones: hace que podamos caer en alucinaciones colectivas.

USANDO NUESTROS RECURSOS PARA PENSAR

Puede resultar natural imaginar que cuando un niño pequeño dice por primera vez «¡perrito!» cuando pasa un perro, comprende que está etiquetando la cosa que ve con una palabra hablada, pero no es necesario. Debemos ser muy cuidadosos en no confundir competencia y comprensión. De hecho, puede que para los adultos sea útil, en términos prácticos (por muy erróneo que sea en términos teóricos), suponer mucha más comprensión en las primeras expresiones de sus hijos de la que realmente tienen, ya que así les construyen un andamio conversacional que ayuda a los niños a acostumbrarse a su lengua (McGeer 2004). De este modo, las protoetiquetas allanan el camino para las etiquetas, y las protopeticiones hacen lo propio con las peticiones. Puede llevar cientos de horas de «conversación» que un niño realmente le coja el truco, y aún

más que un niño comprenda qué es lo que está haciendo cuando tiene una conversación.

Recordemos (de los capítulos 9 y 10) que un meme es una manera, de comportarse o de hacer, interna o externa, que se puede transmitir de anfitrión a anfitrión mediante copias. Dicho de otro modo, tal como veremos más adelante en este capítulo, es como una aplicación de software o un applet, en realidad es muy y muy parecido a una aplicación: una estructura informacional comparativamente simple y diseñada (por la selección natural, en sus primeros días) y que puede añadir talento, un poco de pericia, útil o inútil. Estas aplicaciones suelen ser visibles o audibles a los demás cuando se ejecutan (cuando las palabras se oyen o se expresan, cuando la acción es observada o efectuada) y también son visibles o audibles para uno mismo, durante la autoobservación normal: oyes cómo hablas, ves o sientes cómo realizas un gesto, cómo actúas, cómo te mueves, etc. Pero una cosa es oírte hablar a ti mismo y otra muy diferente es darte cuenta de que estás diciendo algo. Un niño de primaria puede tener un desliz y llamar a su maestra «¡mamá!», con el consiguiente regocijo de la clase… y la vergüenza propia ¡cuando se da cuenta! A veces, también los adultos hablan sin darse cuenta de que están hablando: supervisan sus emisiones verbales, quizá corrigen palabras mal pronunciadas, pero son bastante ajenos a la relevancia de sus acciones, casi como si estuvieran hablando en sueños. Sabemos que se autoobservan porque a veces vuelven sobre sus palabras (demasiado tarde): «¿Qué es lo que acabo de decir? ¿Seguro que he dicho eso?».

Mi propuesta es que debemos imaginar a nuestros ancestros pioneros en el uso del lenguaje parloteando locuazmente, logrando grandes ventajas de estos nuevos hábitos, aunque a veces ahuyentaran a sus presas. A veces apenas conseguían comunicarse, mientras que otras veces lo lograban con elocuencia (pero una elocuencia inconsciente, no preparada), y solo poco a poco llegaron a identificar retrospectivamente lo que habían estado haciendo. Sugiero que, igual que a los antílopes nunca les viene a la cabeza lo que están haciendo cuando se dedican a brincar, tampoco nuestros ancestros eran conscientes de lo que estaban haciendo con las palabras hasta que pasaron de la selección inconsciente, tal como Darwin expresó acertadamente, a la selección metódica de memes plenamente domesticados. Hoy en día, los niños hacen una transición tan rápida del balbuceo al habla y la comprensión que resulta fácil verlo como un cambio súbito (similar al emblemático tropiezo de Helen Keller [1908] con la bomba de agua, una anécdota imborrable,

aunque de fidelidad incomprobable).******** En cambio, es probable que en los primeros días del lenguaje el proceso fuera mucho más lento y menos eficiente, sin beneficiarse aún de los ajustes genéticos que optimizaron la adquisición, por ejemplo potenciando la curiosidad por los sonidos del habla de manera que se dedicara más atención (y esfuerzo cerebral) a su análisis. Me permito especular que hubo un proceso gradual y acumulativo de competencia creciente que llevó a la autoobservación, que, a su vez, llevó a la reflexión, la cual, a su vez, llevó a la aparición de nuevas cosas sobre las que pensar: palabras y otros memes en nuestra imagen manifiesta.

Mi aseveración es que fue esta serie de pequeños pasos lo que alzó las palabras y otros memes hacia nuestra ontología, hacia nuestra imagen manifiesta, abriendo así nuevos y extensos paisajes a nuestra curiosidad inherente y permitiendo el inicio de la exploración «de arriba abajo» del espacio de diseño.[6] En 1991, en mi libro La conciencia explicada presenté una versión simple de esta innovación, en la que imaginaba unos ancestros que, habiendo adquirido el hábito de plantear preguntas simples cuando se topaban con algo, descubrían que, a veces, hallaban las respuestas que estaban buscando incluso cuando no había nadie más para oírlos. Se encontraron, pues, respondiendo a sus propias preguntas. Inventaron el hablarse a uno mismo, lo que tuvo unos beneficios inmediatos, e inmediatamente apreciados. ¿Qué ayuda aporta este proceso de hablarse a uno mismo? ¿Por qué no es como pagarse a uno mismo por hacerse la comida? Este aire de paradoja se desvanece en cuanto identificamos que podemos «saber cosas» en una parte de nuestro cerebro a la cual otras partes no pueden acceder cuando es necesario. La práctica de hablar consigo mismo crea nuevos canales de comunicación que, de vez en cuando, pueden hacer que el conocimiento oculto se manifieste.

«En aquel momento parecía una buena idea» es una frase habitual tras algún tremendo error, pero lejos de ser un distintivo de estupidez, si esta frase de arrepentimiento se expresa con sinceridad, es un emblema de inteligencia; un ente pensante que puede recordar con precisión lo que realmente pensaba y cómo se sentía ya está a más de medio camino de depurar sus propios procesos de pensamiento y, así, de no volver a tropezar con la misma piedra. Cuanto más sutil sea el error, más valioso es el hábito de la autoobservación que se activa cuando alguien se habla a sí mismo de forma regular (la próxima vez que intentéis solucionar un problema, dejaos llevar por un monólogo vocalizado; es una buena forma de detectar lagunas en vuestro pensamiento).

Podría ser posible que los animales sin lenguaje se devanaran los sesos en busca de pistas durmientes, pero las pruebas de un comportamiento oculto de este tipo son más bien escasas; en cualquier caso, nuestra práctica de un autocuestionamiento explícito tiene otra gran ventaja: que nuestras meditaciones sean más fáciles de recordar, y así podemos repasarlas a posteriori. Y una vez que adquirimos la costumbre de situarnos en modo de autocuestionamiento, toda nuestra I+D se vuelve mucho más «de arriba abajo», con búsquedas más dirigidas y confiando menos en las variaciones aleatorias y la conservación. No se elimina la aleatoriedad (o lo que cuenta como tal en la evolución: la separación entre la generación de candidatos y los criterios de selección); incluso los trabajos de investigación más sofisticados a veces se aceleran con episodios deliberados de ensayo y error «al azar». Pero el espacio de búsqueda puede restringirse gracias a la información ya adquirida en otros contextos, que permite descartar grandes zonas como improbables o irrelevantes (pero solo si el pensador puede recordarlo a tiempo). Hablar consigo mismo, plantearse cuestiones o simplemente repetir internamente palabras relevantes (las «palabras clave») es una forma eficiente de sondear las redes de asociación conectadas a cada palabra y de recordar posibilidades pasadas por alto que podrían ser relevantes para la confusión actual.

La invención de la computadora digital por parte de Turing nos puede servir de ejemplo emblemático de un diseño inteligente de arriba abajo. De lo que podemos espigar a partir de los registros de esos años heroicos, su ruta hasta la gloria no fue precisamente directa, sino más bien una exploración tortuosa de posibilidades, desvíos, callejones sin salida y objetivos reajustados, y una buena dosis de ayuda fortuita gracias a encuentros con otros pensadores que meditaban acerca de otros problemas. En el ideal de una mente humana consciente, todo su conocimiento es equitativamente accesible, siempre está disponible cuando se lo necesita, nunca está distorsionado, es un caos de expertos listos para echar una mano a tiempo. El cerebro bayesiano de un animal o un niño está bien pertrechado para hacer parte de este sondeo de arriba abajo de pistas sin el beneficio de recursos de pensamiento más avanzados, pero la mente de un adulto puede (en contadas ocasiones) ejercer una severa disciplina sobre esa tripulación de expertos, priorizando, sofocando competiciones inútiles y organizando equipos de búsqueda.

En su libro sobre la evolución del lenguaje, Hurford, en cierto momento, dice: «Los animales no humanos, con un cierto grado de evolución, tienen ricas vidas mentales, un tipo de inteligencia natural que les permite apañárselas bien con su mundo» (2014, p. 101). La segunda parte de la frase es incontestable, pero todavía no está claro en absoluto si realmente uno tiene que tener una «rica vida mental», en el sentido de una vida consciente como la nuestra, para espabilarse por el mundo. He sostenido que el cerebro puede ser un generador de expectativas bayesiano sin que por ello tengan que poseerse los ofrecimientos rastreados y aprovechados. Si los animales no humanos pueden pensar sobre el pensamiento, o si pueden enfrascarse en estados mentales «meta-» sobre sus propios estados mentales, es una pregunta que aún no tiene respuesta, y una «rica vida mental» sin capacidad para intensificarse mediante la reflexión no es tan rica, en realidad. ¿Cuán valiosa es una percepción exacta y cautivadora de los olores, pongamos por caso, en una mente que no es muy capaz de comparar, recordar, maravillarse, reflexionar y apreciar los resultados de estas reacciones secundarias y terciarias?[7] Cuidado: no estoy diciendo que los animales no humanos y los niños no sean conscientes (he aplazado toda discusión sobre la conciencia). Si es mejor considerar que la conciencia puede darse en diversos grados (tal como he planteado y defenderé otra vez cuando corresponda), podemos concebir diversos tipos de conciencia que pueden ser «ricos» en ciertos sentidos pero no en otros.

La riqueza es un buen concepto para afianzar esta perspectiva. Consideremos el tipo afortunado que posee una parcela de tierra donde está enterrado un tesoro pirata con lingotes de oro. En cierto sentido, es rico, pero si no sabe de la presencia del tesoro, o no lo puede hallar y, por tanto, no puede hacer uso de él, su riqueza es picwickiana como mucho.[8] De modo parecido, antes de que un niño pueda rastrear, apreciar, reflexionar, comparar, evocar, recordar y anticipar las diversas diferencias de su mundo que ahora tiene disponibles (en algún sentido), podríamos hablar de riquezas potenciales y si un animal nunca puede acceder a esta perspectiva reflexiva esas riquezas ni tan solo llegan a ser potenciales (mostrar a tus loros muchos episodios de Cosmos garantiza que ante sus ojos y oídos pase una oleada de información vívida y detallada, pero no esperes que tus loros aprendan ni una pizca de ciencia a partir de esta experiencia).

Aún estoy aplazando un ataque frontal a los enigmas de la conciencia, de modo que tened presente que no estoy afirmando (ni negando) que

rastrear, apreciar, reflexionar, etc. sean procesos necesariamente conscientes. Lo que afirmo es que se trata de importantes competencias de respuesta en ausencia de las cuales la conciencia (sea lo que sea) podría ofrecer bien poca cosa a la competencia (y, por tanto, a la comprensión) en desarrollo de nuestros ancestros. En el capítulo 5 sostuve que la comprensión es una cuestión de grados y que es un producto de la competencia, no una fuente independiente de ella. Ahora podemos ver cómo se crea este producto. Los ofrecimientos básicos, dones de la naturaleza genéticamente transmitidos y potenciados mediante el aprendizaje bayesiano para aprovechar los patrones del entorno, proporcionan un conjunto de competencias para todos los animales locomotores, desde los insectos y los gusanos a los elefantes y los delfines. Algunas de las competencias más impactantes, naturalmente, se consideran evidencias de algún tipo de comprensión. Los talentos de un cuervo, capaz de usar herramientas, o de un oso pardo pendiente de los turistas en el parque de Yellowstone, demuestran grados de adaptación a situaciones únicas que dan pie a asumir comprensión. Incluso podríamos decir que la comprensión, en estos dominios conductuales, es precisamente esto: no un conocimiento teórico comunicable, sino una experiencia práctica.

Si esta experiencia es el límite de la comprensión no humana (hasta donde podemos suponer), ¿qué es lo que añade la comprensión humana a todo esto? No solo la capacidad de comunicar la experiencia (o el conocimiento factual) a los conespecíficos, sino también la capacidad de tratar cualquier tema bajo consideración como una cosa a examinar, analizar y clasificar gracias a nuestra capacidad de representarlo explícitamente mediante palabras, esquemas y otros recursos de la autoestimulación. Solo con el cerebro no se puede pensar mucho, pero con estos sencillos recursos se abre ante nosotros todo un mundo de exploración consciente. Tal como ha pasado muy a menudo antes, una competencia diseñada por selección natural (ahora tanto genética como memética) crea un talento, una competencia sin comprensión (en este caso, el dominio del lenguaje) que, a su vez, abre grados cada vez más altos de competencias adicionales. Se trata de metacompetencias, con las que usamos nuestros recursos de pensamiento para pensar no solo en comida, vivienda, puertas, recipientes, peligros y los demás ofrecimientos de la vida cotidiana, sino también para pensar sobre pensar en comida y vivienda, y para pensar sobre pensar en pensar en comida y vivienda, tal como esta misma frase demuestra.

La filosofía ha sido el principal hogar académico de la metarrepresentación durante milenios. Resulta fascinante y divertido ver cómo Platón, Sócrates y, después, Aristóteles, se esfuerzan con la novedosa tarea de intentar comprenderlo todo, incluyendo el propio proceso de la comprensión. Los pasos reflexivos (explorar el significado del significado y la comprensión de la comprensión, usar palabras para analizar palabras) llevaron finalmente a la aceptación de que, tal como lo expresó Douglas Hofstadter, «cualquier cosa que puedas hacer, yo la puedo metahacer». Esto no siempre conduce a resultados y a veces los participantes pueden quedar atrapados en una sala de los espejos sin relación clara con el mundo real, pero el tremendo poder de un meme (un metameme) da lugar a estos excesos.

LA ERA DEL DISEÑO INTELIGENTE

Este proceso recursivo, acumulando montones y montones de metarrepresentaciones, ha provocado la explosión de MacCready que ha florecido durante los últimos diez o veinte mil años, y aún está acelerando. Antes he sugerido que los memes son parecidos al software; ahora me gustaría ampliar y precisar esta afirmación.

Sopesemos los applets de Java. Es posible que os encontréis con ellos a diario, algo casi seguro si pasáis algo de tiempo en Internet, pero por diseño en general son invisibles. Java es el invento responsable, en gran parte, de aportar a Internet su versatilidad (de James Gosling y otros en 1991); deja descargar pequeños programas (los applets de Java) de sitios web que os permiten hacer crucigramas, resolver sudokus, explorar mapas, ampliar fotos, participar en juegos de aventura con otros jugadores repartidos por todo el mundo y también hacer tareas informáticas «serias». Un diseñador de páginas web puede escribir programas en lenguaje Java sin necesidad de saber si las computadoras de los usuarios que acceden a esa web son Mac, PC o Linux, porque un applet de Java siempre se ejecuta en una máquina virtual Java (JVM, Java Virtual Machine) diseñada específicamente para ejecutarse en máquinas Mac, PC o Linux. En pocos segundos, la JVM se descarga e instala automáticamente en la computadora y, luego, el applet se ejecuta en la JVM, como por arte de magia (quizá de vez en cuando os habréis fijado

en la descarga de actualizaciones de Java en vuestras computadoras… ¡o quizá no! En condiciones ideales, podéis olvidaros de qué JVM está instalada en vuestra computadora y podéis confiar en que a cualquier web que accedáis, los applets pertinentes se ejecutarán en vuestra JVM o, en caso contrario, primero se instalará la actualización adecuada de la JVM sin molestias).

La JVM es la interfaz que compila el código Java en el código que se ejecutará en el hardware concreto que se está usando. El lema de Java es WORA: Write Once, Run Anywhere ('escribe una vez, ejecuta en todas partes'), lo que implica que los problemas de diseño solo tienen que solucionarse una única vez. No cuesta ver la excelencia de este arreglo cuando lo comparamos con la máquina virtual de español instalada en vuestro cerebro (si queréis, para mantener la analogía con la JVM, SVM Spanish Virtual Machine). Obviamente sois hispanohablantes, o no podríais estar leyendo este libro. Cuando escribí este capítulo, tuve muchos problemas a solucionar en términos de exposición y de clarificación de los temas, pero no tuve que preocuparme en absoluto de los detalles neuroanatómicos de los diversos cerebros que podrían leerlo; lo escribí una única vez (bueno, escribí muchos borradores, pero solo he publicado uno) y ahora puede «ejecutarse» en cualquier cerebro que disponga de una SVM. Como el libro estaba originalmente escrito en inglés, solo era accesible a cerebros que dispusieran de una máquina virtual de inglés (o EVM, English Virtual Machine), y por ello era necesaria una traducción para que los cerebros con una SVM en lugar de una EVM pudieran experimentar sus efectos; en caso contrario tendrían que haber aprendido inglés.

Pensemos en lo fácil que resulta para maestros, guías, informadores y formadores (y también para timadores, propagandistas y evangelizadores) diseñar un discurso que tenga buenas posibilidades de instalar memes adicionales en vuestros cerebros, informándoos, influyéndoos y armandoos con nuevos recursos. Me basta con escribir mi invento y cuando vosotros leáis lo que he escrito, estaréis descargando una aplicación en vuestros cerebros, tan simple como eso. Mi libro Bombas de intuición y otras herramientas de pensamiento (2013) consta de 72 capítulos, en cada uno de los cuales presento por lo menos un recurso de pensamiento para que el lector se lo descargue, junto con consejos sobre cómo crear y analizar tales recursos. Sin duda, el libro presupone una buena cantidad de conocimientos generales previos por parte del lector en diversos campos, de modo que tiene que tener aplicaciones ya instaladas

para esos recursos de pensamiento. Escogiendo una página al azar (p. 71), entre las palabras que utilizo ahí sin explicación previa tenemos: margarita, estrella de mar, oso hormiguero, aceitunas, mancha, lengua, herencia, infancia y categoría. Si no tenéis alguno de estos familiares recursos en vuestro equipo, os costará entender la página. Otros términos más difíciles (ontología, Umwelt, imagen manifiesta, umbral de fusión de parpadeo, término masa y clasificador) los explico sobre la marcha: nuevas aplicaciones que dependen de otras aplicaciones nuevas que dependen de aplicaciones más antiguas.

Tal como ha observado Jackendoff (2002), muchos experimentos con animales como sujetos requieren que estos estén entrenados para realizar una tarea muy concreta a cambio de una pequeña recompensa alimenticia, en la mayoría de casos. Si queremos que un gato, un mono o una rata pulse una palanca cada vez que ve un cuadrado en lugar de un triángulo o cada vez que oye «biiiip» en lugar de «beeeep» quizá necesitaremos cientos o incluso miles de ensayos de prueba para que el animal se convierta en un sujeto fiable y atento. Cuando los científicos cognitivos quieren que los participantes humanos ejecuten la misma tarea, unas pocas instrucciones simples en un idioma que los participantes entiendan (incluyendo la promesa de una compensación económica al final de la sesión) y unos pocos ensayos bastan para que la mayoría de voluntarios humanos se conviertan en unos perfectos ejecutores de la tarea durante todo el experimento. Ciertamente, las personas pueden descargar y ejecutar una máquina virtual sin necesidad de pruebas y errores ni de aprendizaje asociativo, asumiendo cientos de papeles diferentes a voluntad, con rapidez y fiabilidad.

Así, la plasticidad neural y los hábitos de atención necesarios para convertirnos en imitadores (básicamente sin comprensión) ofrece la base para la adquisición del lenguaje, lo que, a su vez, fundamenta el virtuosismo cognitivo necesario para escuchar, comprender y seguir instrucciones. La comprensión del lenguaje aporta una comprensión mucho más general del comportamiento y del mundo en su conjunto. Pero nuevamente, no deberíamos exagerar la eficacia de la transferencia que permite el lenguaje. Algunas habilidades son más difíciles de transmitir que otras y algunas exigen un considerable entrenamiento y trucos mnemotécnicos, como vimos en el capítulo 6. Los nombres fantasiosos y formas imaginarias de las constelaciones son una reliquia de los primeros esfuerzos en la transmisión de información, pero no debemos suponer que algún astrónomo ingenioso ideó todo el plan como una mejora

pedagógica consciente; seguramente las asociaciones fueron el resultado de tradiciones acumuladas a lo largo de años y años de cuentos e historias, cuya utilidad solo resultó evidente de forma gradual; en todo este proceso, las asociaciones de segunda fila, menos recordables, se extinguieron, incapaces de reproducirse entre tanta competición, hasta que solo quedaron los campeones mnemónicos.

Por supuesto, el paralelismo entre los applets de Java y los memes (o entre las computadoras digitales y los cerebros) no es exacto, como ya hemos visto, pero dejadme hacer hincapié en algunos aspectos. Es cierto que una computadora sin sistema operativo (Windows, MacOS, Linux, etc.) es casi inútil, y también es cierto que no puedes pensar mucho solo con tu cerebro, por las mismas razones. Pero los cerebros, a diferencia de las computadoras digitales, no tienen un control centralizado y jerárquico, con pulsos de sincronización, policías de tráfico para determinar prioridades, etc. En consecuencia, no me basta con descargar mi aplicación en vuestros cerebros y dejar que se ejecute; el control debe ejercerse mediante la negociación y la diplomacia y, en ocasiones, también mediante súplicas, amenazas u otros empujones emocionales. En general, tendré que garantizar vuestra atención, vuestra cooperación y, en cierto modo, vuestra confianza, porque estáis y debéis estar alerta frente a posibles manipulaciones por parte de otros agentes. Al principio, las computadoras se diseñaban para ser unos ejecutores obedientes de cada una de las tareas que se les planteaban (¡no se admiten preguntas!), pero con el crecimiento de Internet, la cada vez más frecuente presencia de programas malignos (malware) diseñados para apropiarse de computadoras, con propósitos perversos, ha obligado a los diseñadores de los sistemas operativos a incluir capas y capas de protección, de modo que hoy en día, la situación predeterminada cuando llega cualquier software nuevo es NO EJECUTAR ESTE CÓDIGO sin el permiso explícito del usuario o administrador de la computadora. En estos momentos, los usuarios humanos somos el eslabón más débil en la seguridad; cada día llegan correos electrónicos fraudulentos más y más atractivos, diseñados inteligentemente para ganarse nuestra confianza.

Con ayuda de la figura 13.1 podemos representar cómo se acumula la cultura humana a un ritmo cada vez más rápido, ya que el aumento de los procesos de búsqueda dirigida y de planteamiento de problemas de arriba abajo lleva a una resolución de problemas más eficiente. Entre las innovaciones capacitadoras tenemos «invenciones» tan brillantes como la escritura, la aritmética, el dinero, los relojes o los calendarios, cada una de

las cuales ha aportado un sistema de representación nuevo y provechoso que ha dotado a nuestra imagen manifiesta de cosas con las cuales hacer cosas más portátiles, independientes, manipulables, identificables y recordables, las cuales nos permiten aprovechar nuestro creciente dominio de otras cosas. Por lo que podemos decir, estas invenciones fueron «darwinianas», es decir, invenciones sin inventores o autores previsores, más parecidas a las alas de un pájaro que a los álabes de una turbina.

Las motivaciones flotantes de los rasgos y estructuras de estas invenciones han sido captadas, representadas y celebradas gradualmente por sus beneficiarios, ingenieros inversos retrospectivos que pudieron explicar al mundo la utilidad concreta de la representación fonémica de las palabras, del cero como número, de las monedas a prueba de falsificaciones, de la representación del tiempo con una línea, un círculo o un volumen mediante un ciclo breve y fijo de nombres para los días. Todos estos artefactos transmitidos culturalmente, abstractos o concretos, son, sin duda alguna, recursos bien diseñados de pensamiento, pero no son la creación de diseñadores inteligentes individuales.[9]

Esto también es cierto de muchos recursos (quizá la mayoría) para otros propósitos que hemos ido añadiendo a nuestras cajas de herramientas durante milenios. En The Secret of Our Success (2015), el antropólogo Joseph Henrich explora con bastante detalle artefactos extraordinarios, como los cayacs, y sostiene que es muy poco probable que ningún diseñador/artesano sea responsable de algo más que un puñado de características o de unas pocas mejoras de los artefactos en los que se inspiraba. En la mayoría de casos, la sabiduría popular que se acumula alrededor de estos artefactos es el producto de una teorización retrospectiva, no de una invención original. Probablemente está formada, en buena parte, por subproductos de los intentos de pasar las habilidades a los aprendices, comentarios de segunda mano que pueden mejorar la comprensión de todo el proceso; o puede ser una «falsa conciencia», «teorías» persuasivas pero erróneas sobre el tema en cuestión.

La adopción y el uso de nuevos memes para solucionar problemas suele generar una mayor comprensión en los que se están trabajando la solución, pero no siempre. A veces, al «solucionar» un problema se llega a la solución sin querer, o con una comprensión errónea de lo que se ha hecho; las personas que lo han logrado propagan sus triunfos y, a menudo, su reputación. Si alguna vez has hecho una jugada «brillante» en el

ajedrez pero solo te has dado cuenta de su brillantez a posteriori, y te lo has callado, ya sabes de qué va la cosa. Hay una altruista tendencia general a atribuir a los innovadores más comprensión previa de la que realmente tenían, lo que ayuda a perpetuar el mito de las capacidades casi divinas de nuestros genios más famosos y, por extensión, de todos nosotros. ¿Acaso no somos los animales racionales homenajeados desde Aristóteles? La misma ilusión de comprensión que tiende a nublar nuestra apreciación de las astutas tretas de los animales (y de las plantas y de las bacterias, como hemos visto) distorsiona nuestras valoraciones de nuestros colegas humanos… y de nosotros mismos.

Una de las reglas generales de una estafa bien diseñada es que debe aprovecharse de nuestro amor propio como inteligentes detectores de fraudes, dándonos oportunidades de «demostrar» el poder de nuestro escepticismo antes de poner a prueba nuestro dictamen. A menudo quedamos estupefactos al descubrir lo vulnerables que podemos ser a estafas y fraudes y en el propio reconocimiento de nuestro daño también solemos cometer el mismo pecado por segunda vez, atribuyendo más inteligencia de la que merecen a los timadores que nos han engañado; al fin y al cabo, sus trucos suelen ser antiguos, comprobados y refinados por generaciones de estafadores previos antes de ser transmitidos. Un caso clásico es el trile, que se remonta, por lo menos, a la antigua Grecia. Se muestran tres cubiletes boca abajo, con una bolita debajo de uno de ellos; el timador desplaza rápidamente los tres cubiletes por la mesa y pregunta «¿dónde está la bolita? ¡Hagan sus apuestas!». Es posible que algu-

na vez os hayáis topado con un grupo de gente apiñada entorno a un trilero; con prudencia os acercáis, protegiendo con fuerza la cartera de vuestro bolsillo, y observáis durante un rato. Algunos tipos ganan, la mayoría pierden, pero en estos casos descubrís el juego de manos del trilero: «¡ya lo tengo!». Tras valorar la situación, os lanzáis: sacáis vuestra cartera y observáis con agudeza los movimientos subrepticios que habíais identificado antes. Apostáis… ¡y perdéis! ¿Cómo puede ser? Sencillo: hay dos juegos de manos: el que el trilero os ha permitido detectar y rastrear y el que continúa oculto a vuestra mirada, casi invisible para todo el mundo excepto para los magos entrenados y los estafadores. Además, la muchedumbre agolpada a vuestro alrededor está formada básicamente por cómplices del trilero, ganchos que representan bien el papel de espectador algo bobo, poco atento y demasiado confiado, al cual podéis «explicar» los errores que ha cometido, como demostraréis cuando entréis en la partida. El trilero tiene un talento auténtico, sabe cómo hacer los

movimientos clave y sabe cuánto aguantar esperando que os decidáis a apostar, pero el truco es nuevo solo para vosotros.

En círculos sociales más prestigiosos, la misma aura idealizada de comprensión tiñe nuestra percepción en campos como la agricultura, el comercio, la política, la música y el arte, la religión, el humor… en resumen, en toda la civilización humana. Los granjeros que obtienen las mejores cosechas deben saber lo que están haciendo, como los corredores de bolsa más exitosos, los músicos más populares o los políticos más reelegidos. Aunque aceptamos que la suerte tiene un papel difícil de detectar, seguimos la regla general de buscar el consejo de aquellos que tienen éxito (o parecen tenerlo). Los modelos económicos clásicos, con su suposición de agentes racionales que emplean la información disponible para optimizar sus decisiones de mercado, muestran el innegable atractivo de la suposición de racionalidad. Al idealizar y simplificar en demasía las interacciones reales que se dan en el mundo, convierten la embrollada actividad humana en un conjunto de fenómenos más o menos predecible y más o menos explicable, cosa que funciona bastante bien para muchos propósitos. Quizá no somos perfectamente racionales, pero tampoco somos tan irracionales. La explicación de la ley de la oferta y la demanda, por ejemplo, es transparente: como, en general, los compradores intentarán minimizar sus gastos y los vendedores intentarán maximizar sus cobros netos, los precios de cualquier producto subirán y bajarán en función de su disponibilidad y deseabilidad. No se trata de una ley de la física; solo es una prescripción del comportamiento racional, pero seríamos ingenuos de ir contra la ley de la oferta y la demanda (si tenemos conocimiento interno de algo relevante que el mercado no comparte, parecería que engañamos a la ley de la oferta y la demanda, pero en realidad nos estamos aprovechando de ella: cualquier ventaja que obtengamos estará dictada por la amplia tendencia de nuestras víctimas a comportarse como prescribe la ley).

Cuando echamos un vistazo a las descripciones tradicionales de la evolución cultural, en el sentido neutro de cambio cultural a lo largo del tiempo, el tema dominante es el modelo económico, como si toda la evolución cultural humana hubiera tenido lugar cerca de la cumbre de la comprensión y todos los objetos implicados se valorasen por buenas razones. En este modelo, los elementos culturales constan de bienes, que se adquieren racionalmente, se conservan y se reparan y luego se legan a la siguiente generación, junto con la experiencia necesaria para preservarlos y disfrutarlos. Protegemos lo que valoramos, lo transmitimos

a aquellos cuyos intereses deseamos mejorar e invertimos en las infraestructuras necesarias para mantener los bienes en buen estado. Este modelo no sirve para la chatarra cultural (los hábitos despreciables pero infecciosos que se propagan a pesar de que los valoramos negativamente)[10] ni para las variantes culturales furtivas que se reproducen sin que nos demos cuenta (el cambio de significado de enervar, por ejemplo), pero funciona bastante bien para las partes de la cultura humana que algunos miembros de la sociedad estiman valiosas. La ópera se mantiene gracias a patrocinadores, que también donan a conservatorios de música y conceden premios por logradas interpretaciones de los elementos legados. Los museos, las bibliotecas, las universidades, el patrimonio arquitectónico, los monumentos y campos de batalla famosos requieren un costoso mantenimiento, y los que valoran estos elementos se esfuerzan para garantizar que haya una dotación económica suficiente para su conservación. Muchas personas dedican su vida a conservar artesanías y actividades tradicionales: la tejeduría, el bordado, el encaje de bolillos, las muñeiras, el vals vienés, la doma o la ceremonia japonesa del té. Las religiones invierten buena parte de las donaciones que reciben de sus feligreses a conservar, modernizar y amueblar sus edificios y pagar a su personal, a menudo dejando una cantidad más bien módica para las tareas de beneficencia.

Cuando examinamos con detenimiento las alturas de la cultura humana contemporánea, con sus magníficos artefactos, tanto abstractos como concretos, que aseguran nuestra salud y seguridad, nos liberan de pesadas cargas y llenan nuestro recién conseguido tiempo libre con todo tipo de arte, música, ocio y aventura, hallamos una abundancia de ejemplos de entidades inteligentemente diseñadas. Sabemos que han sido inteligentemente diseñadas porque hemos presenciado y registrado los procesos muchas veces y hemos hablado con los diseñadores, prestando atención a sus explicaciones y justificaciones y aprendiendo sus técnicas y objetivos, sus criterios estéticos y, en muchos casos, comprobando sus demostraciones y sus documentaciones.

Sin duda, vivimos en una era de diseño inteligente, que se remonta a varios milenios (tan atrás como la documentación disponible). Los constructores de las pirámides sabían lo que estaban haciendo y habían articulado planes y objetivos que comprendían y ejecutaban con precisión, organizando miles de trabajadores humanos en un proceso muy distinto al de la construcción de los castillos de las termitas: se basaba en un control de arriba abajo y en un enorme grado de comprensión. Esto no

significa que los constructores de pirámides no se basaran en una inmensa cantidad de experiencia, memes que habían sido perfeccionados y optimizados gracias a una replicación diferencial relativamente ciega durante los milenios anteriores. Cada generación tiene un legado de conocimiento acumulado la mayoría del cual, hoy en día, ha quedado confirmado casi por completo en miles de aplicaciones, incluyendo aquí rotundos fracasos por parte de rebeldes que decidieron volverse contra la tradición y probar otros métodos y normas. Aunque buena parte de este conocimiento conseguido con grandes esfuerzos también se ha explicado, expresado, analizado y racionalizado en tratados sobre todos los temas imaginables, no deberíamos caer en el error de pensar que los autores de estos tratados eran los inventores o diseñadores de los principios y prácticas que enseñaban. En general, desde la época de Aristóteles hasta la actualidad, las explicaciones y justificaciones de nuestro almacén de conocimiento general son algo así como historia whig, escrita por los vencedores, que en tono triunfal explican los descubrimientos y pasan por alto los costosos errores y las búsquedas equivocadas.

Si queremos tener una idea más clara del diseño inteligente, debemos resistirnos a esta mitología retrospectiva y prestar algo de atención a la historia de los fracasos y a todos los diseños de segunda fila que podemos identificar a nuestro alrededor. En primer lugar, realza nuestra apreciación de las mejores obras. Por ejemplo, mi comprensión de la grandeza del arte se ha visto ampliada gracias a horas y horas de visita en museos de arte de segunda categoría por Europa, donde se pueden ver centenares de obras menores del mismo estilo y con las mismas aspiraciones que Rafael, Rembrandt, Rubens o Rodin y apreciar inmediatamente cuán inanes y descompensadas son en comparación con las obras maestras. Las bibliotecas musicales de todo el mundo contienen partituras de miles de sinfonías clásicas que seguramente nunca se volverán a interpretar, y por buenas razones (siempre puede haber unas pocas obras maestras inesperadas entre todas esas mediocres composiciones, pero pocos músicos tienen ganas de enfrascarse en la tediosa tarea de explorar en busca de estas joyas). Esta acumulación cultural casi invisible tiene un papel no muy diferente al de todos los organismos de todas las especies que murieron sin descendencia pero cuya competición templó el carácter de aquellos cuyos descendientes están vivos en la actualidad. Haydn, Mozart y Beethoven tuvieron que cultivar sus talentos en un mundo en que había muchos compositores solo un poco menos talentosos o quizá solo algo desafortunados.[11]

En el transcurso de toda esta cría deliberada de memes domesticados hay mucho de toma de decisiones consciente y concienzuda y el modelo económico explica más o menos bien el perfil de decisiones en estos campos, tal como lo hace en ámbitos más comerciales. La ley de la oferta y la demanda dirige las trayectorias futuras tanto de las academias de doma como de los fabricantes de piruletas. Pero incluso en las instituciones más burocratizadas y racionalizadas hay patrones de cambio (de evolución) que se resisten a una explicación mediante el modelo económico y que se muestran como un simple ruido o una casualidad. Algunas de estas escandalosas variaciones sin duda son ruido, no mejores que la mera aleatoriedad, pero parte de ellas no están dirigidas por las presiones racionales del mercado, sino por la carrera armamentística de la evolución memética darwiniana: hay una incesante competición entre memes para la replicación, y explicar las preferencias que alimentan la competición nos exige descender de las alturas de la racionalidad pura (v. figura 13.2) y echar un vistazo al punto medio. Es aquí donde agentes con semicomprensión se enfrascan en proyectos más o menos bien diseñados que generan un montón de debilidades, nuevos objetivos que atraen a los explotadores de debilidades, que, a su vez, también son agentes con semicomprensión que se dan cuenta, con diversos grados de exactitud, que lo mejor para sus intereses es adoptar memes, adaptar memes y revisar memes diseñados (ahora gracias a una mezcla de selección natural y diseño inteligente) para aprovecharse de las flaquezas que les ha parecido detectar en su presa. Estas carreras armamentísticas son la última oleada de las creativas carreras armamentísticas que han impulsado la evolución genética durante tres mil millones de años, y se diferencian, sobre todo, en que implican un importante grado de comprensión para engrasar y acelerar las innovaciones y las reacciones. Esta comprensión es, de forma abrumadora, producto del lenguaje y de otros medios de comunicación.

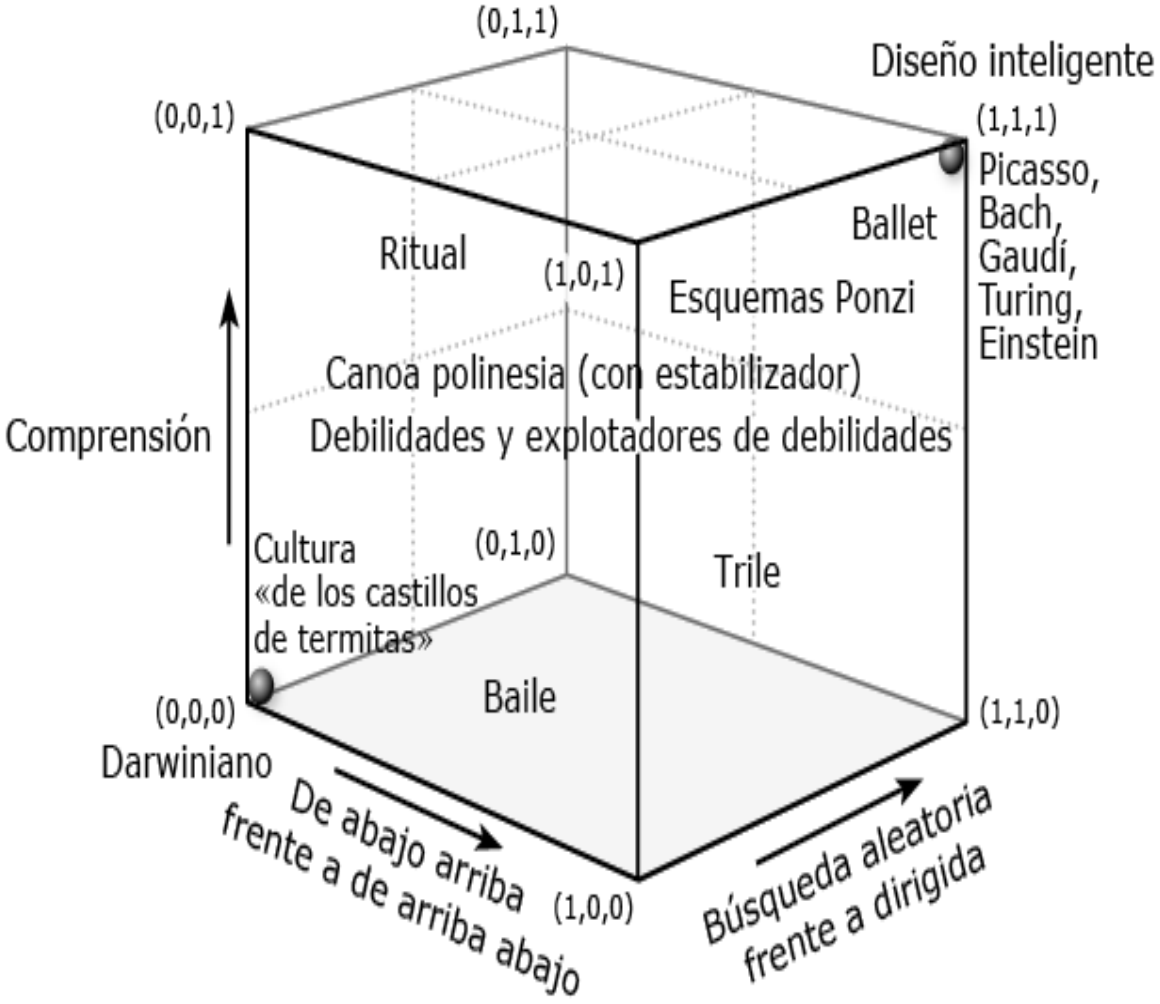


FIGURA 13.2. Espacio darwiniano de la evolución cultural, con fenómenos intermedios

Al añadir al escenario estas variedades de intercambio de información, se modifica el entorno selectivo de un modo revolucionario. La evolución darwiniana normal por selección natural, al carecer de toda previsión, circula por poblaciones que también carecen de previsión. Los supervivientes son afortunados de estar mejor pertrechados para las nuevas condiciones que el resto de su generación, y no hay ningún tipo de aviso, o es muy escaso, de los nuevos riesgos y oportunidades. Si un ñu descubre un nuevo movimiento evasivo que le salva de los depredadores, no hay ninguna manera rápida de que los otros ñus de la manada lo adopten, y sin duda, no se podrán propagar «rumores» sobre el tema por las diversas manadas de la sabana. En nuestro caso, hay tradiciones bien

establecidas de obtener y transmitir nuevos consejos en cuanto están disponibles (sean buenos o malos). Todas las fábulas, cuentos populares y remedios secretos, toda la sabiduría del mundo que se adquiere «en el regazo de la madre» se complementa con chismorreos y rumores y se perfecciona gracias a la escritura, la impresión y todos los medios actuales de alta tecnología. Esto acelera e intensifica las carreras armamentísticas y las complica con subtramas y opciones secundarias a medida que las personas equipan sus imaginaciones con toda la información y la usan para imaginar variaciones.

Hoy en día es más difícil hallar «señales» de la estafa del príncipe nigeriano que hace veinte años, porque es un timo bien conocido y puede servir como tema humorístico con seguridad (Hurley, Dennett y Adams 2011). Ahora bien, ¿cómo es que la estafa aún persiste y por qué los timadores continúan empleando la misma historia sospechosa de un príncipe nigeriano que te quiere transferir una fortuna a tu cuenta y necesita tu ayuda? Porque los estafadores no tienen intención de gastar su tiempo y sus esfuerzos en personas astutas. El coste de enviar millones de anzuelos iniciales en forma de correos electrónicos es trivial comparado con el coste (en tiempo y esfuerzo) de halar cuando algún incauto ha mordido el anzuelo. No te puedes permitir atraer demasiadas presas. La mayoría de víctimas potenciales que han picado pronto sospecharán algo y puede que descubran la verdad, abandonando antes de haber pagado nada, e incluso puede que algunos se diviertan a costa de los estafadores durante algún tiempo. Para minimizar el coste de halar peces que luego escaparán del anzuelo antes de haberlos capturado, el señuelo inicial está diseñado para ser tan ridículo que solo las personas más crédulas, simplonas y sin preocupaciones podrían creérselo. En resumen, la estafa incorpora un filtro para evitar a los escépticos, de modo que el esfuerzo se pueda concentrar en las víctimas más vulnerables.

Los consejeros políticos y los publicistas, analizadores de tendencias y especuladores, observan agresivamente y perturban la memosfera, explorando nuevas acciones, nuevas oportunidades, nuevas fisuras en el blindaje de escepticismo y prudencia que todos empleamos a diario para protegernos, excepto las personas más ingenuas. Todo el mundo quiere «volverse viral» con sus mensajes favoritos, a la vez que busca nuevas formas de ignorar el exceso de captadores de atención que los asalta. Esta fluidez de la transmisión de información en la cultura humana, y su uso para combatir, desacreditar, descartar y también perfeccionar, mejorar,

adaptar y propagar nuevos memes deja a la evolución memética darwiniana en segundo plano.

¿Acabo de admitir, acaso, que la memética no puede ser un recurso teórico valioso para modelizar la evolución cultural actual (y futura)? Recordemos otra vez la afirmación de Steven Pinker, citada en el capítulo 11:

Los rasgos más destacados de los productos culturales, a saber, su ingenio, belleza y verdad (análogos al complejo diseño adaptativo de los organismos) provienen de los cálculos mentales que «dirigen» —es decir, inventan— las «mutaciones» y que «adquieren» —es decir, entienden y comprenden— los «caracteres». (1997)

Esto se acerca bastante a la verdad para algunos productos culturales, pero, como hemos visto, al insistir en que la invención y la comprensión son los únicos fenómenos sobre el tapete, Pinker limita nuestra atención a los tesoros culturales, solamente, y exagera el papel del diseño inteligente en su creación. Además, Pinker subestima la distorsión incorporada en nuestras prácticas lingüísticas: una demanda de razones que supera nuestra capacidad de satisfacerla. Cuando adquirimos nuestra lengua materna, se nos introduce en el espacio de razones; ya no se trata de motivaciones flotantes, sino de motivaciones ancladas, razones representadas que poseemos y respaldamos, nuestras respuestas a todos los por qué. La transición es incómoda, porque no siempre tenemos buenas respuestas, pero ello no es óbice para que sigamos jugando la partida.

— ¿Por qué construyes las barcas de esta forma?

— Porque es como las hemos construido siempre.

— Pero ¿por qué es esto una buena razón?

— ¡Porque lo es!

La normatividad de la aportación de razones se impone incluso cuando no tenemos respuestas; hay una obligación de tener razones que puedas ofrecer para explicar tu comportamiento. Estamos ante el enfoque intencional, la actitud intencional, en plena acción, siempre asumiendo racionalidad: ¡tiene que haber una buena razón, o de lo contrario no lo estaríamos haciendo así! Ante esta exigencia, la tentación, en la que caemos a menudo, es coger una razón de la nada y plantearla, con un aire de complicidad, lo que genera un práctico pedazo de ideología que incluso podría ser correcto. Si no se nos ocurre ninguna razón plausible, siempre se puede recurrir a la sapiencia de nuestros ancestros:

Sabían las razones, y debemos agradecer que nos enseñaran, aunque no nos dijeran por qué.

Afirmación que está a un paso de

Los caminos del Señor son inescrutables.[12]

Los niños, antes de disponer del lenguaje, deben entrenarse de un modo no muy diferente a un perro. «¡No toques! ¡No! ¡Ven! ¡Quieto! ¡Muy bien! ¡Otra vez!» Los padres no suponen que los niños entenderán por qué; solo son niños. Pero cuando empiezan a hablar, empezamos a ofrecerles razones: «¡No toques; caliente!» «¡No toques; sucio!» «Come; así crecerás». La obediencia, incluso la obediencia ciega, es útil como base; luego ya habrá tiempo para explicar y argumentar. «Porque lo digo yo» es un paso importante. Más adelante, al crecer, se nos presentan las

normas de la sociedad humana, que incluyen de forma destacada la suposición de racionalidad en los otros, especialmente en los adultos (pero no, claro está, en nuestros animales domesticados). Un caballo vaquero que muestre perspicacia en las tareas que realiza con su jinete nos maravilla y nos sorprende… ¡qué inteligente! Los perros pastores son ejemplos apasionantes, y son apasionantes porque son excepcionales. No asumimos los mismos raseros para los animales que los que asumimos para nosotros mismos.

Estamos ante una característica del uso del lenguaje que es imposible de eliminar; la suposición de comprensión está predeterminada, y tanto es así que, cuando se viola, nos quedamos confundidos. Estás en un país extranjero y tienes que preguntar una dirección, con optimismo pruebas con tu lengua materna, por si, por casualidad, tu interlocutor la conoce; el nativo sonríe y escucha, asintiendo con alegría, de modo que prosigues lleno de confianza, pero cuando has acabado te das cuenta de que tu interlocutor no ha entendido ni una palabra. ¡Vaya chasco!

Vivimos nuestra cotidianeidad inmersos en la suposición de comprensión, tanto en los extraños con los que nos topamos como en nuestros amigos y familiares. Esperamos que las personas esperen que tenemos razones, razones que podemos expresar, para lo que sea que estamos intentando hacer, y siempre inventamos razones sin esfuerzo alguno (a menudo sin darnos cuenta) cuando nos preguntan y no tenemos una respuesta preparada. Es ante este trasfondo que la memética parece tan equivocada y tan subversiva, como afirma Pinker. ¡Nosotros somos los razonadores! ¡Los diseñadores inteligentes! Somos los explicadores de todo. Es una afrenta sugerir que algunas de nuestras brillantes ideas no sean más que perturbaciones cerebrales provocadas por memes invasores luchando por el dominio. ¡Nosotros tenemos el control!

Y sí, tenemos el control… hasta cierto punto, hasta un punto que nos resulta satisfactorio. Este es el triunfo de la invasión de los memes: ha convertido nuestros cerebros en mentes (nuestras mentes) capaces de aceptar y rechazar las ideas que encontramos, de descartarlas o desarrollarlas por razones que solemos poder expresar, gracias a las aplicaciones que tenemos instaladas tras nuestro cogote.

Pero no siempre tenemos razones, lo que nos distingue de los agentes racionales concebibles pero imposibles de la GOFAI, el sueño definitivo del diseño inteligente. En estos sistemas de conocimiento, las razones se

representan «hasta el fondo», en el sentido de que todo el pensamiento se considera una especie de demostración de teoremas. Cuando una GOFAI os dice algo, en principio siempre podéis pedirle una razón y esta debería estar disponible, porque sea lo que sea que la GOFAI os haya dicho, tuvo que estar generado a partir de una inferencia racional de su base de datos de axiomas. Os puede enseñar la demostración y vosotros podréis criticarla paso por paso, si queréis.

La llegada del «aprendizaje profundo» y de los métodos bayesianos ha sido recibida con sentimientos contrastados por muchos científicos cognitivos. ¿Por qué? El hecho de que estas nuevas estructuras cognitivas funcionen tan bien es sorprendente y maravilloso y sus aplicaciones están llamadas a comerse el mundo, pero… aunque nos aportarán muchas respuestas a preguntas difíciles, no podrán decirnos por qué. Serán oráculos que no podrán participar en el juego de aportar razones (tema sobre el que volveremos en el último capítulo).

PINKER, WILDE, EDISON Y FRANKENSTEIN

En esta vida, la primera obligación es ser totalmente artificial.

O

SCAR WILDE

En un lúcido comentario (como es habitual) sobre mis conferencias de 2009 del departamento Mente-Cerebro-Comportamiento de la Universidad de Harvard, Steven Pinker amplió su crítica a los memes:

El diseño sin diseñador es crucial para la evolución biológica, pero es perverso para la evolución cultural: en realidad sí que hay un diseñador, el cerebro humano, y no hay nada místico ni misterioso en decirlo. (23 de

abril de 2009, disponible en línea en https://www.youtube.com/watch?v=3H8i5x-jcew)

Pinker y yo estamos de acuerdo en que la evolución biológica (genética) es diseño sin diseñador, y también en que las facultades del cerebro humano como diseñador dependen de haber sido bien diseñado sin diseñador (hasta hace muy poco). Sin embargo, lo que argumento es que buena parte de esta I+D es, y tiene que ser, evolución memética, no genética. Mucho antes de tener cerebros diseñadores, teníamos cerebros que adquirían diseños sin diseñador en forma de memes invasores.

Pero los memes pueden efectuar mucho más trabajo explicativo incluso estando ya bien inmersos en la era del diseño inteligente. En su comentario, Pinker presenta dos buenos ejemplos: el origen de la palabra acné, que procede de un error de transcripción del término griego acmé, que significaba 'culminación', y que comparaba con la acuñación del término inglés palimony ('pensión alimenticia tras la separación de una pareja no casada legalmente') creado en un «arrebato de creación, de perspicacia o de ingenio», una palabra compuesta por pal ('compañero') y alimony ('pensión alimenticia'). Pero no sabemos si esto es cierto. El término fue usado por primera vez en público por el abogado matrimonialista Marvin Mitchelson en 1977, cuando su clienta, Michelle Triola Marvin presentó una demanda fallida contra el actor Lee Marvin, con quien no había estado casada legalmente (véase el artículo «palimony» en la Wikipedia en inglés), pero no sabemos si la creación de Mitchelson fue un golpe de genio o el resultado de una tediosa búsqueda de variantes («alimony, ballimony (mmmm…), calimony, dallymony (quizá)… galimony (ejem)… palimony… ¡eh! Igual puedo usar esta»).******** Conozco a dos empedernidos verbívoros que me han confesado que esto es lo que hacen, obsesivamente, con casi cada frase que oyen, descartando el 99% o más de los juegos que sopesan rápidamente. No se trata de un diseño muy inteligente, con toda esta búsqueda semidirigida, laboriosa y promiscua… ¿o sí?

Un rasgo curioso de nuestra apreciación del ingenio y la inteligencia es que preferimos no saber cómo se logra. Consideramos a Oscar Wilde como una persona muy ingeniosa, pero indudablemente su reputación quedaría en entredicho si supiéramos que se pasaba despierto muchas noches cavilando obsesivamente: «¿Qué podría contestar si alguien me

pregunta…? ¿Cuál sería mi sucinta opinión sobre…?». Supongamos que hemos descubierto que, con paciencia, determinó y perfeccionó varias docenas de comentarios agudos cada noche, clasificándolos ingeniosamente en su memoria, listos para lanzarlos en cuanto la situación lo exigiera, en un abrir y cerrar de ojos, pues la brevedad es la base del ingenio. Continuarían siendo sus ocurrencias, y su gracia dependería del momento. El momento es importante para casi cualquier acto inteligente, y es por eso por lo que resulta inteligente anticipar y resolver con antelación los problemas siempre que sea posible. La estrategia de fuerza bruta de producción de ocurrencias de Wilde nos puede decepcionar, pero nos obliga a fijarnos en el hecho de que todas las respuestas inteligentes dependen de una I+D costosa y que no supone gran diferencia cómo se distribuye el trabajo a lo largo del tiempo, siempre que los efectos se den en el momento adecuado.

La famosa cita de Thomas Alva Edison que dice que la genialidad es un 1% de inspiración y un 99% de transpiración provoca en muchas personas la antipática idea de que esto quizá sea cierto para Edison y otros pocos inventores obstinados pero poco brillantes, pero que los genios de verdad invierten los porcentajes (¡tal como alardeaba Picasso!). Esta actitud nubla nuestra percepción de la relación entre la creatividad y la conciencia, una de las distorsiones de la gravedad cartesiana, que empezamos a sentir cada vez con más fuerza a medida que nos acercamos a la conciencia. Queremos que nuestras mentes sean «inspiradas» y «asombrosas» y tendemos a ver los intentos de diseccionarlas como asaltos intolerables (y, para colmo, incultos) contra nuestros dones más preciados. «Cualquiera que sea lo bastante ingenuo como para creer que nuestras mentes son solo un montón de maquinaria neuronal debe de tener una familiaridad patéticamente pobre con la magia de nuestras mentes.»

Lee Siegel (1991), en su notable libro sobre la magia callejera de la India, tiene un fragmento que he citado a menudo:

Explico que «estoy escribiendo un libro sobre magia» y me preguntan: «¿La magia de verdad?». Con magia de verdad la gente quiere decir milagros, actos taumatúrgicos y poderes sobrenaturales. «No», respondo, «trucos de magia, no magia de verdad». Así, la magia de verdad se refiere

a la magia que no es real, mientras que la magia que es real, la que realmente se puede hacer, no se considera magia de verdad. (p. 425)

Incluso muchos neurocientíficos, naturalistas entregados, tienden a estremecerse cuando empiezan a acercarse al tema de cómo la maquinaria neuronal obra sus maravillas y, con una atractiva pizca de modestia, niegan que estén siquiera intentando abordar el «Problema Difícil» (Chalmers 1995), dejando la «magia de verdad» de la conciencia como un misterio cósmico para los siglos venideros, cuando pueda resolverse con ayuda de revoluciones de la física hoy en día inimaginables. Los que piensan de este modo deberían reflexionar sobre un secreto del oficio divulgado por uno de los mejores prestidigitadores, Jamy Ian Swiss (2007): «nadie pensaría que nos esforzaríamos tantísimo para engañarte. Esto es un secreto y un método de la magia». Si nunca intentas imaginar cómo se podría hacer, con una enorme cantidad de trabajo, siempre te quedarás perplejo. En el capítulo 2 ya dije que los magos no tienen reparos; los cerebros, tampoco.

Como no hay nada que sea «magia de verdad», cualquier camino que siga la mente de un genio, al final tiene que explicarse mediante la cascada de grúas que han sido diseñadas y construidas durante los últimos miles de millones de años. De un modo u otro, la mente fue alzada (gracias a un buen diseño) hasta una zona del espacio de diseño desde la cual sus posteriores incursiones fueron admirablemente rápidas y eficaces. Tanto si la «inspiración» es un 1% como un 99%, tiene que ser un rasgo de un vehículo que permite explorar el espacio de diseño y que tiene su propia historia de I+D, en absoluto sobrenatural; quizá una enrevesada combinación de genes, educación, experiencia vital, orientación y quién sabe qué más… la dieta, haber oído casualmente una melodía, desequilibrios emocionales temporales o permanentes y quizá algo de «locura», algún trastorno o alteración mental que tiene beneficios fortuitos (beneficios para el mundo del arte, si no para el artista).

Cuando intentamos determinar cuánta invención podemos atribuir a la creatividad de un artista, podemos obtener una idea aproximada imaginando las zonas concretas del espacio de diseño atravesadas para llegar al producto final. Un experimento mental nos ayudará a centrar la situación:

Supongamos que el doctor Frankenstein diseña y construye un monstruo, Spakesheare, que inmediatamente después se sienta a escribir una obra de teatro, Spamlet. ¿Quién es el autor de Spamlet?

En primer lugar, quiero hacer hincapié en lo que considero que es irrelevante en este experimento mental: no he dicho si Spakesheare es un robot, construido con metal y chips de silicio o construido a partir de tejidos humanos, o células o proteínas o aminoácidos o átomos de carbono. Mientras el diseño funcione y la construcción sea obra del doctor Frankenstein, no tiene ninguna relevancia de qué materiales esté hecho. Podría ser que la única manera de construir un robot lo bastante pequeño y rápido y que no consumiera demasiada energía para sentarse en un banquillo y escribir una obra de teatro fuera a partir de células artificiales llenas de hermosas proteínas motoras y otros nanorrobots de carbono. Se trata de un problema científico y técnico interesante, pero sin relevancia para nuestro experimento. Por la misma razón, si Spakesheare fuera un robot de metal y silicio quizá tendría que ser tan grande como una galaxia, si es lo que se necesita para obtener la complejidad necesaria en su programación (y tendríamos que anular el límite de la velocidad de la luz en beneficio de nuestro experimento mental). Como este tipo de limitaciones técnicas no suelen tenerse en cuenta en los experimentos mentales, que así sea. Si el doctor Frankenstein decide hacer su IA robótica a partir de proteínas y cosas parecidas, es su problema. Si su robot puede aparearse con los seres humanos y, en consecuencia, crear lo que podríamos decir que es una nueva especie, es un tema fascinante, pero lo que aquí nos preocupa es la supuesta creación de Spakesheare, Spamlet. Repitamos la pregunta:

¿Quién es el autor de Spamlet?

Para poder captar la cuestión, debemos mirar hacia el interior y observar qué es lo que sucede en Spakesheare. En un caso extremo, nos encontramos con un archivo (si Spakesheare es un robot con una memoria

informática) o una versión memorizada de Spamlet, listo para ejecutarse. En este caso extremo, sin duda el doctor Frankenstein es el autor de Spamlet, que utiliza a Spakesheare como creación intermedia, un mero dispositivo de almacenamiento y reproducción, un procesador de texto muy sofisticado. Todo el trabajo de I+D ya se había realizado previamente y se había copiado en Spakesheare, de un modo u otro.

Podemos visualizar el proceso con más claridad si imaginamos un subespacio del espacio de diseño, que denominaré la Biblioteca de Babel, en honor al clásico cuento de Jorge Luis Borges del mismo título (1941). En este cuento, Borges nos invitaba a imaginar un almacén lleno de libros que, a ojos de sus habitantes, es infinito; finalmente deciden que no es así, pero que podría serlo, pues parece que en sus anaqueles (sin orden alguno, desgraciadamente) hay todos los libros posibles. Ahora ya estamos preparados para volver a esa aguja en un pajar, Spamlet, y considerar qué trayectoria se siguió hasta este lugar concreto de la Biblioteca de Babel en la historia real. Si hallamos que el viaje entero había concluido en el momento en que se construyó y se llenó de información la memoria de Spakesheare, sabremos que Spakesheare no tuvo papel alguno en la búsqueda. Empezando por el final, si encontramos que el único papel de Spakesheare era pasar el texto almacenado por un

corrector ortográfico antes de usarlo para dirigir sus movimientos de escritura, no nos impresionará demasiado la afirmación de la autoría spakeshearíana. Se trata de una parte medible, pero Vastamente pequeña, de la I+D total. Hay una pléyade de textos casi gemelos de Spamlet, más o menos cien millones de diferentes mutantes menores que tienen una única errata sin corregir, y si ampliamos muestro horizonte para incluir una errata por página, ya nos acercamos al dominio de un número Vasto de variaciones sobre un mismo tema. Yendo un poco más atrás, cuando pasamos de las erratas a los thinkos, ya estaremos en el terreno de la verdadera autoría, a diferencia de la mera edición. La relativa trivialidad de la corrección y edición pero, a la vez, su innegable relevancia para conformar el producto final queda bien demostrada en términos de nuestra metáfora del espacio de diseño, en que cada pequeño proceso de alzamiento sirve para algo, y a veces unos pequeños desplazamientos hacia arriba pueden dirigirnos hacia una trayectoria totalmente nueva. En este punto, podemos citar a Mies van der Rohe, como viene siendo habitual: «Dios está en los detalles».

Ahora demos la vuelta a nuestro experimento, tal como Douglas Hofstadter ha recomendado (1981) y observemos el otro extremo, en que el doctor Frankenstein deja casi toda la tarea en manos de Spakesheare. La situación más realista sería que Frankenstein ha equipado a Spakesheare con un pasado virtual, una reserva de experiencias vitales pseudomemorizadas, en las cuales basarse para satisfacer su pasión obsesiva, instalada por Frankenstein, de escribir una obra de teatro. Podemos suponer que entre estos pseudorrecuerdos hay muchas tardes en el teatro, muchos libros leídos y también algunos amores no correspondidos, ciertas decisiones difíciles, unas cuantas traiciones vergonzosas y otras experiencias parecidas. ¿Qué sucede a continuación? Quizá algún fragmento de una historia de «interés humano» en las noticias será el catalizador que incitará a Spakesheare a un frenesí de creación y prueba, rebuscando en su memoria detalles y temas útiles, transformando (transponiendo, mutando) lo que encuentra, sacudiendo las piezas en estructuras temporales y prometedoras que compiten para su finalización, pero la mayoría de las cuales quedarán descartadas por el corrosivo proceso de crítica que revela fragmentos útiles de vez en cuando. Toda esta búsqueda a varios niveles estaría guiada, de algún modo, por evaluaciones generadas internamente a diversos niveles, incluyendo evaluaciones de evaluaciones de… los productos de las búsquedas en curso. Si el doctor Frankenstein hubiera previsto realmente toda esta actividad hasta su escala más pequeña, hasta el nivel más turbulento y caótico, y hubiera diseñado a mano el pasado virtual de Spakesheare y toda su maquinaria de búsqueda para producir solo este producto final, la obra Spamlet, entonces el doctor Frankenstein sería, una vez más, el autor de Spamlet, pero también sería, para decirlo rápido, Dios; una presciencia tan Vasta sería algo milagroso. Pero recobrando un poco de realismo en nuestra fantasía, podemos situarnos en posiciones bastante menos extremas y suponer que el doctor Frankenstein no pudo prever todo esto en detalle, sino que delegó en Spakesheare buena parte de la ardua tarea de completar la trayectoria en el espacio de diseño hasta llegar a una obra literaria, a determinar mediante una I+D posterior que se daría dentro del mismo Spakesheare. Sin más que colocarnos en esta posición hemos llegado a las cercanías de la realidad, pues ya disponemos de ejemplos reales de impactantes autores artificiales que superan Vastamente la previsión de sus propios creadores. Nadie ha creado todavía una obra de teatro artificial digna de una mínima atención, pero un jugador de ajedrez artificial (el Deep Blue de IBM) y un compositor artificial (el EMI de David Cope) han logrado resultados que, en algunos

aspectos, igualan a los mejores productos que el genio humano puede concebir.

Podemos adaptar el experimento mental de la trayectoria de Spakesheare por el espacio de diseño para analizar otros campos de la creatividad y el descubrimiento humanos. La sofisticación del texto de Spamlet se puede medir con precisión en algunas dimensiones, porque todas sus posibles variantes están almacenadas en la Biblioteca de Babel de Borges y se pueden identificar y evaluar sus diferencias (erratas y thinkos y aburridos errores de digresión). Podemos identificar los plagios culpables y la inspiración laudable, aunque no hay una frontera clara que separe estos dos reinos. Ahora que disponemos de impresoras 3D, tenemos un montón de formatos para representar cualquier objeto tridimensional (un broche, un abrelatas eléctrico, una escultura de Rodin, un violín Stradivarius) como un archivo digital que se puede comparar con sus variantes. Cuando Picasso ve el sillín y el manillar de una bicicleta como lo que necesita para crear la cabeza de un toro, no está plagiando, sino que echa mano de algunas formas y materiales cuidadosamente diseñadas y optimizadas para sus propios propósitos, una práctica artística que dispone de un término específico: objets trouvés.

Así, todos los inventos y apropiaciones concretas (herramientas, instrumentos, objetos artísticos, máquinas) se pueden «situar» de forma unívoca en un espacio multidimensional, la Biblioteca de Objetos, si queremos llamarlo así, y podemos representar en él sus linajes de descubrimiento o invención, como los linajes en el árbol de la vida, pero con una buena cantidad de anastomosis. ¿Cuán similares deben ser dos cosas, y en qué dimensiones, para que cuenten como un único diseño digno de ser protegido mediante una patente o unos derechos de autor? No podemos esperar encontrar respuestas objetivas sencillas a tales cuestiones. Las invenciones abstractas, en particular, desafían una colocación «en orden alfabético» en cualquier espacio de cualquier dimensionalidad, lo que constituye un revigorizante aviso a todos los antirreduccionistas, pero la transmisión de invenciones abstractas de una mente a otra sigue basándose en alguna combinación de palabras e imágenes u otras representaciones que se puedan grabar y transmitir. La brillante imagen o fábula de Turing de una máquina avanzando sobre una cinta de papel, leyendo y borrando (o no) una marca y luego siguiendo adelante, no es un rasgo opcional de su invención. Esto nos aporta otra perspectiva sobre el requisito de la ley de que las patentes tengan una utilidad demostrada y de que una innovación musical tiene que tener

alguna «expresión fijada» para poder someterse a derechos de autor (véase el apartado «Secretos comerciales» del capítulo 6). El espacio de diseño multidimensional en que ha crecido el árbol de la vida ha generado un espacio de diseño hijo para nuestras creaciones, y tiene más dimensiones y más posibilidades que su progenitor; y nosotros somos la única especie (hasta el momento) capaz de explorarlo.

Steven Pinker está en lo cierto cuando dice que «el cerebro humano es un diseñador», pero esto no se debe considerar como una alternativa al enfoque memético, sino como una continuación del enfoque memético hacia la era del diseño semiinteligente gradualmente desdarwinizado. Nuestra visión tradicional del genio lo representa como algo totalmente diferente de la selección natural por lo que respecta a sus poderes creativos, y no es por casualidad que los genios se consideren muchas veces como divinos y sobrenaturales. Al fin y al cabo, hemos creado a Dios a nuestra imagen (manifiesta) y semejanza, una prolongación del enfoque intencional al cosmos. Pinker, un darwiniano y naturalista incondicional e imaginativo, ha hecho lo que ha podido para arrancar el misterio de nuestra concepción de la mente (1997) y ha recibido ataques furibundos por sus intrépidas incursiones darwinianas en las humanidades, pero aquí, la gravedad cartesiana le desvía ligeramente del rumbo que nos llevará de las bacterias a Bach.

BACH COMO HITO DEL DISEÑO INTELIGENTE

Está en perfecto acuerdo con el plan de la naturaleza, determinado por la selección natural, que la materia excretada para liberar al sistema de sustancias superfluas o nocivas deba ser empleada para [otros] propósitos de gran utilidad.

C

HARLES DARWIN, 1862

Juzgar un poema es como juzgar un pastel o una máquina. Se le pide que funcione. Es solo porque un artefacto funciona que podemos inferir la intención de un artífice.

W. W

IMSATT Y M. BEARDSLEY, 1954

Turing, Picasso, Einstein, Shakespeare, todos ellos son ejemplos excelentes de diseñadores inteligentes, pero me gustaría centrarme un poco más en Johann Sebastian Bach (1685-1750), que combina las virtudes de los diseñadores inteligentes de un modo que resulta muy informativo. Por un lado, nació en una familia de gran tradición musical (su padre y sus tíos eran músicos profesionales y, de sus veinte hijos, cuatro de los que sobrevivieron hasta la edad adulta fueron también compositores de renombre), con lo que hay buenas posibilidades de que tuviera una dotación inusualmente potente de «genes de la música» (en este punto, no deberíamos ignorar el papel de la cultura al explicar que la música «viene de familia»; si te crías en una familia musical, sean cuales sean tus genes, estarás rodeado de música, participarás en la vida musical de la familia y estarás imbuido con la importancia de la música). Por otro lado, complementó su competencia musical «divina» con exhaustivos estudios; Bach fue un experto del contrapunto y la armonía, y estaba muy bien informado acerca de los compositores más antiguos, cuyas obras analizó en profundidad. También era algo así como un manitas del siglo XVIII, gran organista y experto también en el diseño, reparación y mantenimiento de estas maravillas mecánicas, respetados antecesores de los sintetizadores electrónicos de la actualidad.

Podemos comparar Bach con otro gran compositor, Irving Berlin, incapaz de leer partituras, que solo podía tocar el piano en una única tonalidad (fa sostenido mayor, de modo que podía tocar casi todo solo con las teclas negras del piano) y que tenía que confiar en un «secretario musical» para escribir y armonizar sus melodías. Parece ser que ejercía un control muy estricto sobre las armonizaciones que escribía su ayudante, pero lo hacía siempre «de oído», sin conocimiento alguno de la teoría de la armonía. Cuando ya era rico y famoso (cosa que sucedió muy pronto en su carrera), se hizo construir un piano especial, con un teclado deslizante, para así

poder tocar siempre en fa sostenido, desplazando el teclado de modo que transportara las notas a otra tonalidad.

¿Puede ser brillante un compositor así, sin formación alguna? Por supuesto. Entre sus más fervientes admiradores como compositor estaban Cole Porter y George Gershwin, ambos músicos con una profunda formación. Gershwin, que no era ningún advenedizo, calificó a Berlin como «el mejor autor de canciones que haya existido nunca» (Wyatt y Johnson 2004, p. 117). Berlin no ha sido el único gran compositor incapaz de leer música. The Beatles seguramente son el ejemplo más destacado, con varias docenas de canciones que luego han inspirado versiones, adaptaciones y homenajes en todo el espectro musical, desde el hip-hop a los cuartetos de cuerda.

Leonard Bernstein, compositor y director de orquesta con una formación clásica y alumno de Harvard, publicó un ensayo en 1955 con el divertido título «Why don't you run upstairs and write a nice Gershwin song?» ('¿Por qué no te vas arriba y compones una bonita canción al estilo de Gershwin?') en el que explicaba su fracasado intento, en colaboración con un musicalmente refinado amigo, de componer una canción en el estilo del Tin Pan Alley.******** Nadie era más consciente del abismo que hay entre las habilidades técnicas y el éxito (económico y musical) que el propio Bernstein, que finalizaba su ensayo con una melancólica confesión: «La verdad es que sería bonito oír a alguien silbando algo mío, en algún sitio, solo por una vez». Dos años después, su deseo se hizo realidad con West Side Story, pero Bernstein, a pesar de su genialidad, no fue una sensación de la noche a la mañana, como tampoco lo fue Bach. A pesar de la enorme producción musical de Bach como compositor, cuando murió era conocido básicamente como un gran organista, y tuvieron que pasar más de cincuenta años para que otros compositores, críticos y músicos empezaran a cimentar su fama y provocar la fijación de sus memes en la historia de la música occidental.

El fracaso de las composiciones de Bach para propagarse como un incendio no se debía a ningún desinterés de Bach para asegurar su rápida replicación. De hecho, fue un maestro en el delicado arte de inspirarse en la tradición para lograr éxitos. En Leipzig fue nombrado maestro de capilla de la iglesia de Santo Tomás y de otras iglesias de la ciudad, cargo que ocupó durante más de veinticinco años, hasta su muerte en 1750. Una de sus tareas era la de componer e interpretar cantatas, elaboradas piezas musicales para coro y solistas, acompañados por una pequeña orquesta o,

a veces, solo por el órgano. Compuso centenares de cantatas, una para cada domingo del calendario litúrgico y adaptada a la celebración y las lecturas de ese día. Parece que compuso cinco ciclos completos (Jahrgänge) de cantatas, de los cuales solo se conservan tres (así, un centenar de sus cantatas se debe haber perdido para siempre).

A partir de su segundo año, adoptó la política de basar cada cantata en un coral, un himno luterano preexistente, que sin duda sería familiar para la mayoría de feligreses. Estas melodías habían demostrado su capacidad de supervivencia durante décadas o siglos, estando bien adaptadas a los oídos y las mentes de sus anfitriones. Así pues, Bach empezó con buen pie el proyecto de encontrar memes musicales con una buena capacidad de persistencia e incluso de volverse virales. Si te enfrentaras a la tarea de componer cincuenta temas musicales fácilmente tarareables en un año, podrías seguir el ejemplo de Bach y echar mano, por ejemplo, de «Noche de paz», «Jingle Bells», «Oh Susanna», «You Are My Sunshine» (compuesta por Jimmie Davies) y «My Heart Will Go On» (compuesta por James Horner). Bach tomó las melodías de los corales como esqueleto sobre el que construir una música exquisitamente bella y les infundió una nueva vida. Bach era el doctor Frankenstein de la reencarnación musical.

Sin duda, algunos melómanos, al leer estos pocos párrafos sobre Bach, Bernstein, Berlin y The Beatles, estarán a un tris de ver confirmado su prejuicio: los que se aventuran en el tema de los memes son unos incultos sin remedio, incapaces de diferenciar la calidad de la cantidad. En realidad, espero haber provocado exactamente este sentimiento, para desentrañarlo y acabar con él, de una vez por todas. No tiene por qué haber ningún conflicto entre el estudio científico y objetivo de la cultura, de la calidad que sea, y la articulación de opiniones estéticas. Los memetistas no tienen prohibido entrar en el terreno de juego de las opiniones críticas solo porque insistan en medir, predecir y explicar la reproducción diferencial de los memes.

Para equilibrar la ecuación, los expertos en artes y humanidades que tienden a considerar automáticamente que los superventas, las novelas gráficas y los álbumes de platino son una basura artística, deberían deshacerse de tales prejuicios y aprender a percibir la considerable interrelación entre la calidad y la cantidad en todas las artes. Podría añadir que los científicos que desprecian los intentos de explicar la ciencia al público general deberían tomar en cuenta la misma recomendación. En varias ocasiones he sondeado grandes audiencias de científicos que me

han confirmado que las obras de Stephen Hawking, E. O. Wilson, Richard Dawkins, Douglas Hofstadter, Steven Pinker y otros excelentes divulgadores científicos les han servido de inspiración. Es una lástima que desde las artes y las humanidades no hayan surgido más divulgadores equivalentes. Leonard Bernstein lo fue, como sir Kenneth Clark, pero eso fue hace más de medio siglo. Desde entonces, ¿quién ha siquiera intentado hacer trabajos de calidad suficiente para llegar a una gran audiencia? Me vienen a la cabeza Wynton Marsalis y Stephen Greenblatt. Quizá habría menos aspavientos acerca de los desesperados apuros de las artes y las humanidades si en estos campos se dedicara más diseño inteligente a propagar y divulgar sus conocimientos.

No debemos olvidar que es un hecho cultural objetivo que hoy en día hay más personas que pueden tararear «Campana sobre campana» que el coral «Jesus bleibet meine Freude». En el otro extremo, Brahms no solo fue muy popular en vida, sino que también fue el primer artista famoso en grabar una interpretación (un colega de Edison grabó un cilindro con Brahms interpretando un fragmento de su primera danza húngara en 1889). No hay que igualar cantidad y calidad, pero al fin y al cabo el éxito en la propagación es tan necesario para los memes (por muy excelsos que sean) como para los organismos. La mayoría de organismos no dejan descendientes, y la mayoría de libros publicados tienen docenas de lectores, no miles, antes de quedar descatalogados para siempre. Incluso las obras más excelsas de los genios tienen que pasar la prueba de la replicación diferencial. Hoy en día, Moby Dick de Hermann Melville (1851) se considera, con justicia, una de las grandes novelas en lengua inglesa, pero en vida del autor se agotó y no se reeditó, quedando en el olvido hasta 1919, cuando el centenario del nacimiento de Melville dio ocasión a una cierta revisión retrospectiva favorable. A ello siguieron dos versiones cinematográficas (que se alejaban radicalmente de la novela), la edición de Chicago de 1930, con las inolvidables xilografías de Rockwell Kent y, finalmente, la edición de Modern Library Giant de 1943, que también incluía las ilustraciones de Kent. A partir de ese momento, la «inmortalidad» de Moby Dick quedó garantizada.

Quizá, tal como han planteado algunos críticos, las obras de arte verdaderamente grandes siempre viajan bien por el tiempo y el espacio y, de algún modo, encuentran la manera de tocar la fibra sensible de las personas, por muy alejadas que estén de las situaciones en que se crearon. O quizá no. Los procesos con los cuales se recuperan «obras maestras olvidadas» a menudo implican sorprendentes combinaciones de factores

que poco tienen que ver con la brillantez atentamente diseñada por sus autores y más con «tocar la fibra sensible» en los encuentros fortuitos que se suceden. En definitiva, incluso si nuestros cerebros son, en el mejor de los casos, diseñadores inteligentes, la explicación de los patrones culturales dominantes exige separar las circunstancias de la creación de las vicisitudes de la supervivencia.

En el capítulo 2 planteamos y aplazamos una pregunta: ¿por qué ha habido tan pocas mujeres genios famosas? ¿Se trata de los genes, de los memes o de una mezcla de ambos? Nuestra privilegiada perspectiva actual nos sugiere que la respuesta estará más en las características culturales que en la corteza cerebral. Eso no significa que estemos apoyando el desacreditado mantra de los sesenta de que los chicos y las chicas son «biológicamente» iguales y que todas las diferencias son debidas a la socialización y otras presiones culturales. Esto son tonterías políticamente correctas; los cerebros de hombre y mujer no son exactamente iguales. ¿Cómo podría ser así, teniendo en cuenta las diferencias en los papeles biológicos de machos y hembras? Hay docenas de diferencias detectables con facilidad en la neuroanatomía, el equilibrio hormonal y otros signos fisiológicos, cuyos orígenes genéticos no están en duda. Además, estas distinciones físicas llevan a diferencias estadísticamente significativas en los estilos y las competencias cognitivas y emocionales (véase, p. ej., Baron-Cohen 2003). Pero también hay gran cantidad de variación entre mujeres y entre hombres, de modo que algunas mujeres destacan en tareas en que la mayoría de hombres también destacan, y viceversa.

Pero nos estamos preguntando qué es lo que explica la fama para más diseñadores inteligentes hombres que mujeres, y acabamos de ver que esta variable tiene una relación muy débil y poco fiable con la calidad de la mente del diseñador o diseñadora. Lo que hace que una canción o un chiste se vuelvan virales y lo que hace que el nombre y la reputación se vuelvan virales es igual de inescrutable, con una plétora de dignos candidatos de todo tipo de entre los cuales la selección natural escogerá, ciegamente, cuáles amplificar. Esto no significa que sea solo la pura suerte la que determina quién se hace famoso; el desequilibrio entre hombres y mujeres nos garantiza, casi con toda seguridad, que no es así (¿puede haber un gen de la suerte del que las mujeres carecen? Seguro que no). En consecuencia, tiene que haber un cierto desequilibrio en la línea de salida, en que los hombres reciben más números que las mujeres en la lotería de los genios famosos. Si echamos una mirada a los últimos

siglos, la explicación parece evidente: durante milenios, muy pocas mujeres han tenido la más mínima oportunidad de desarrollar sus talentos; es más, incluso no hace tanto, tenemos muchos episodios de científicas que merecían buena parte del crédito de descubrimientos y demostraciones que han otorgado fama a sus colegas masculinos. Y ahora que las mujeres por fin empiezan a hacer valiosos progresos en campos dominados por los hombres, la ciencia se está desplazando hacia equipos de investigadores cada vez más grandes, en que las oportunidades de sobresalir quedan limitadas por las exigencias administrativas de los proyectos. Y sucede lo mismo en otros ámbitos del diseño inteligente. ¿Dónde están los Babbages, Edisons y Watts de la actualidad? Seguramente trabajando en equipos en Google, Apple o Amazon, quizá con un prestigio reconocido entre sus colegas, pero en un relativo anonimato para el mundo. Tal como veremos en el capítulo 15, la heroica era del diseño inteligente empieza a desvanecerse justo cuando sus potenciales heroínas comienzan a demostrar sus talentos.

LA EVOLUCIÓN DEL ENTORNO SELECTIVO PARA LA CULTURA HUMANA

En los albores de la cultura humana, nuestros ancestros albergaban memes beneficiosos sin tener más comprensión de ellos que la que tenían de sus instintos transmitidos genéticamente. No necesitaban comprensión para adquirir nuevas competencias y tampoco obtenían mucha comprensión por tenerlas. La diferencia significativa que aportó la cultura en esos primeros días estribaba solo en que estas maneras (de actuar) transmitidas culturalmente se podían remodelar en una sola generación de anfitriones, transmitirse entre no familiares y adquirirse en vida, no solo en el nacimiento. A medida que los memes se acumularon y se hicieron más y más eficaces para habitar en sus anfitriones (aportando más ventajas o menos inconvenientes, o subyugándolos para su propio beneficio), la imagen manifiesta se fue poblando de más y más ofrecimientos, más y más oportunidades a rastrear, más y más cosas con las que hacer cosas, más y más cosas (palabras) para usar como recursos para ayudar a rastrear cosas, etc. Algunos memes eran recursos, otros eran juguetes, algunos eran distracciones, unos cuantos eran parásitos

incapacitantes. Y todos dependían de la replicación cultural para sobrevivir.

Este entorno competitivo provocó carreras armamentísticas evolutivas con la misma sobreabundancia inevitable de tecnología y contratecnología que encontramos en la guerra humana: mentiras, amenazas, faroles, sondeos y pruebas, como la prueba del shibboleth, por ejemplo, que se puede usar como un pasaporte lingüístico (¿puedes pronunciar shibboleth correctamente? Si no puedes, no eres uno de los nuestros).******** Los sistemas de prácticas recomendadas (trueques, promesas, información mutua, ataques con aviso previo, por ejemplo) pueden demostrar su valía o, de todas formas, fijarse localmente, creando tradiciones conocidas por todos, y que se sabe que son conocidas por todos, lo que los convierte en parte de un entorno conductual más o menos fijo en el cual pueden plantearse, discutirse, aceptarse y rechazarse propuestas. En la actualidad, nadie sabe cuánto de todo esto podría

lograrse sin un lenguaje gramatical plenamente desarrollado, pero sin necesidad de situar todos los elementos en el orden correcto de aparición histórica, aún podemos ver cómo puede surgir gradualmente una sofisticación creciente mediante pequeñas innovaciones, ajustes y perfeccionamientos, formando hábitos e instituciones culturales con un diseño excelente pero sin mucha ayuda del diseño inteligente. «Primero, el aprendizaje; luego, las adaptaciones para el aprendizaje» (Sterelny 2012, p. 25). En cuanto quedó clara la importancia de un rápido aprendizaje, gracias al relativo éxito de los individuos que aprendían más rápido, pudieron empezar a evolucionar maneras de acelerar el proceso, tanto culturales como genéticas. Una de las innovaciones más valiosas fue la práctica de poner marcas o señales en el entorno, para descargar la memoria personal, lo que constituye una de las primeras incursiones de la «mente extendida» (Clark y Chalmers 1998). Estas señales evolucionaron hacia sistemas de numeración y de escritura, que mejoraron el poder de la enseñanza discursiva y, así, en pocos milenios llegamos a Sócrates, Platón y Aristóteles hablando sobre el hablar, pensando sobre el pensar, imaginando repúblicas y teorizando sobre la tragedia y la comedia. La era del diseño inteligente avanza a toda marcha. Las criaturas skinnerianas o popperianas no pueden seguir el ritmo de las criaturas gregorianas, cuyas mentes rebosan de nuevas herramientas y recursos para efectuar evaluaciones aún más rápidas y precisas del complejo entorno que las rodea. Ya no basta con los métodos de fuerza bruta de prueba y error; para poder competir, hay que comprender.

Ahora vivimos en un mundo en que memes sin autores definidos aún se vuelven virales, como los caprichos y las modas, modismos y cambios de pronunciación, pero tienen que competir con invenciones diseñadas meméticamente, creadas con previsión y propósito por forjadores de memes profesionales que representan un papel destacado en la sociedad contemporánea, como escritores, artistas, historiadores o guionistas. Pinker acierta al decir que las creaciones exitosas de estos seres humanos están diseñadas inteligentemente, sin ningún tipo de misterio, pero estas creaciones también nadan en un mar de competidores diseñados semiinteligentemente, diseñados hemisemiinteligentemente y diseñados evolutivamente, y todos dependen de su capacidad para subirse a nuevos cerebros humanos y, así, continuar sus linajes. Pero esto puede cambiar. La evolución cultural ha sido desdarwinizada por sus propios productos, pero su ascendencia darwiniana todavía es bien evidente y los memes sinantrópicos y anónimos, al igual que las bacterias que nos superan en número y en peso, aún nos rodean día tras día.

******** Helen Keller fue una escritora y activista sordociega estadounidense. El

autor se refiere al celebrado momento en que Keller, ante una bomba de agua, comprende por primera vez una palabra, agua, deletreada sobre su mano por su profesora Anne Sullivan. (N. del t.)

******** Todos estos términos inventados son compuestos a partir de alimony (‘pensión alimenticia’): con bally (‘maldito’), con dally (‘flirteo’ o ‘pérdida de tiempo’) con gal (‘chica’) o gall (‘bilis’), hasta llegar al definitivo palimony, con pal (‘compañero’). (N. del t.)

******** Tin Pan Alley es el nombre que recibía un grupo de editores y compositores musicales de Nueva York que dominaron el panorama musical de Estados Unidos durante las décadas a caballo de los siglos XIX y XX. (N. del t.)

******** Lógicamente, para cada lengua, o mejor aún, para cada conflicto entre hablantes distintos, los shibboleth serán diferentes. El artículo «List of shibboleths» de la Wikipedia inglesa presenta ejemplos para muchos pares de lenguas. (N. del t.)

TERCERA PARTE

NUESTRAS MENTES VUELTAS DEL REVÉS

14

LA CONCIENCIA COMO UNA EVOLUCIONADA ILUSIÓN DE USUARIO

[1]

UNA MENTE ABIERTA ACERCA DE LAS MENTES

Por fin estamos listos para juntar todas las piezas y examinar la conciencia humana como un sistema de máquinas virtuales que han evolucionado, memética y genéticamente, para tener papeles muy especializados en el «nicho cognitivo» que nuestros ancestros han construido a lo largo de milenios. Estamos preparados para enfrentarnos sin ambages a la gravedad cartesiana y abordar algunas cuestiones importantes:

¿Cómo logran los cerebros humanos una comprensión «global» usando competencias «locales» y sin recurrir a un diseñador inteligente?

¿Son nuestras mentes diferentes de las mentes de otros animales? Y si es así, ¿cómo y por qué?

¿Cómo se nos hizo manifiesta nuestra imagen manifiesta?

¿Por qué experimentamos las cosas del modo en que lo hacemos?

Hagamos un breve repaso: la evolución ha dotado a todos los seres vivos con los recursos para responder adecuadamente a sus ofrecimientos

particulares, detectando y evitando lo malo, detectando y obteniendo lo bueno, empleando lo localmente útil e ignorando todo lo demás. Esto genera competencia sin comprensión a todas las escalas, desde la molecular hasta la más alta. Dado que puede haber competencia sin comprensión y dado que la comprensión (la comprensión «real») resulta cara, la naturaleza hace un uso intensivo del principio de mínimo conocimiento y diseña criaturas muy capaces, expertas e incluso astutas que no tienen la más mínima idea de lo que hacen ni de por qué lo hacen. Las razones abundan, pero la mayoría son motivaciones flotantes, inimaginables para todos aquellos que se benefician de ellas. Como expertos en ingeniería inversa, podemos determinar la ontología de ofrecimientos en el Umwelt de los árboles, las pulgas y los osos pardos a la vez que nos mantenemos en un agnosticismo total acerca de si hay algo que sea «ser como ellos».

Puede haber razones por las que los árboles hacen cosas sin poseer tales razones («en mente»). ¿Hay algo que una pulga pueda hacer por algunas razones que demuestre que, a diferencia del árbol, posee esas razones, que de algún modo «aprecia» las razones que gobiernan sus acciones? Podría ser que no haya nada que fuera ser como una pulga, sentirse como una pulga, no más que no hay nada que es ser como un ascensor automático. Y ya que estamos en ello, ¿qué es lo que nos hace estar tan seguros de que ser un oso pardo es ser como algo? Sin duda, a nosotros nos parece que tiene que haber algo que sea ser como un oso pardo, sentirse un oso pardo ¿no es cierto? (Basta con mirarlos y oírlos; parece más evidente que hay algo que es ser como un oso pardo que no como una pulga que se esconde en el pelaje de ese oso). Pero quizá nuestra imaginación nos está jugando una mala pasada. Sabemos que hay algo que es ser como nosotros por la simple razón que hablamos de ello cada día, en manifestaciones, quejas, descripciones, poemas, novelas, libros de filosofía e incluso artículos científicos con revisión por expertos. Esto es un rasgo clave de nuestra imagen manifiesta, y este hecho objetivo sería evidente a cualquier científico «marciano» que nos estudiara lo bastante para aprender nuestras lenguas. Nuestras divagaciones introspectivas son comportamientos que son tan observables y medibles como nuestros actos de comer, correr, luchar o amar. ¿Hay algo que nosotros podamos hacer (aparte de hablar sobre ello) que otros animales también puedan y que zanje el tema sobre si tienen una conciencia más o menos parecida a la nuestra? Si los científicos marcianos formulan la cuestión de si otros terrícolas (delfines, chimpancés, perros, loros, peces)

son como los terrícolas parlantes, ¿qué les llamaría la atención? ¿en qué se fijarían y qué les sorprendería? ¿por qué? No estamos solo ante una pregunta científica legítima, estamos ante una pregunta obligatoria, pero es una pregunta que los pensadores suelen disculparse por no considerar, argumentando algo como:

No tengo idea de dónde trazar la línea en la escala de complejidad de la vida. ¿Son conscientes los gusanos, los peces, los reptiles, las aves? Acaso nunca lo sabremos, pero sí sabemos que los seres humanos no somos los únicos conscientes. Es evidente.

Esta respuesta no es aceptable, por dos razones. En primer lugar, la idea de que existe, y tiene que existir, una línea clara de separación, incluso sin saber dónde trazarla, es profundamente predarwiniana; puede haber toda clase de variaciones desde los poetas a los protozoos, pasando por los platelmintos, los perezosos, los pavos y las percas, sin que haya ninguna «esencia» de la conciencia por descubrir. Que la famosa formulación de Nagel (1974) «¿ser como qué?» se trate ahora como unas muletas, como un empujón sin contenido pero que se supone que nos orienta hacia la distinción buscada (y cósmica), debería verse como un bochorno, no como una estrategia aceptable para ganar tiempo (ni, como para algunos filósofos, como un elemento básico de la teoría). En segundo lugar, nos ata de manos con una llamada a la intuición (no es nada más) que podría ser errónea. El agnosticismo temporal sobre la conciencia es bueno (yo mismo lo he propugnado), pero no el agnosticismo cargado con la condición de «por supuesto que hay otros animales que son conscientes, aunque no podamos saber qué significa esto». En el mejor de los casos, se trata de una expresión de confianza en el triunfo final de la imagen manifiesta sobre la imagen científica, a pesar de una larga historia de derrotas. Al fin y al cabo, también parecía evidente que el Sol giraba alrededor de la Tierra. Las personas que ni tan siquiera se aventuran a pensar si los osos pardos son «conscientes como nosotros» (sea lo que sea lo que eso signifique) están sucumbiendo a la ideología, no al sentido común. Sus motivos pueden ser perfectamente honestos (están ansiosos por ampliar el círculo de los seres acreditados con consideración moral porque pueden sufrir), pero hasta que podamos

identificar los rasgos relevantes, los rasgos que importan y explican por qué, esta postura es peor que nada. Aún peor, porque aplaza indefinidamente el planteamiento de preguntas difíciles e importantes sobre qué significa exactamente sufrir, por ejemplo, y si los insectos, los peces o las margaritas pueden sufrir. Por razones morales tenemos que «trazar una línea» y muchas personas (pero no Jains, por ejemplo) se sienten cómodas con las políticas que instan imperiosamente a matar mosquitos, garrapatas, moscas sesé y Plasmodium falciparum (el microorganismo que causa la malaria… ¿puede sufrir, un protozoo?). La mayoría acepta que puedan exterminarse las ratas, pero no sus parientes roedores las ardillas, que un ocurrente humorista denominó «ratas con buenas relaciones públicas». Pero la ciencia debería ser neutral; siempre nos puede sorprender con excepciones importantes a la sabiduría popular.

¿Sentís el tirón de la gravedad cartesiana, ahora? «¿No hay nada que sea ser como un oso pardo, sentirse un oso pardo? ¿estás de broma?» No, en absoluto. Lo que hago es colocar el peso de la demostración en aquellos que quieren discutir que se produce algún cambio especial en algún punto del espectro aparentemente continuo de destreza creciente que colocaría a los árboles y las pulgas (o a las pulgas y los osos pardos, escoged lo que más os guste) en lados diferentes de una gran divisoria cósmica. Puede que haya tal frontera, pero a no ser que «ser como algo» permita a los organismos de un lado hacer algo importante (que podría ser sufrir, aunque deberíamos tener un modo objetivo de detectarlo) que no podrían hacer si estuvieran en el otro lado, se tratará de una frontera solo fundamentada en la tradición popular. No niego que pueda existir tal frontera, pero aplazo el tema para ver hasta dónde podemos llegar sin postularla, que es la forma en que se debería conducir la investigación científica. Si no podéis soportar esta imparcialidad, es que estáis compensando en exceso los efectos de la gravedad cartesiana e inhabilitándoos para la investigación (vale la pena recordar que Descartes solucionó el problema por decreto: solo los humanos eran conscientes; los animales eran meros autómatas). Si rehuimos su veredicto, en el ínterin aún tenemos que trazar la línea moral en algún sitio y mejor que pequemos por exceso, pero a no ser que pongamos en suspensión el criterio científico hasta que tengamos una mejor idea de qué es lo que estamos valorando, no tenemos ningún fundamento para confirmar o desplazar la línea cuando aprendamos más cosas. En el Reino Unido, desde 1986, la ley establece que el pulpo (pero solo

Octopus vulgaris, ningún otro cefalópodo) es un «vertebrado honorífico» que recibe protección legal; así, podéis meter en agua hirviendo una langosta viva, un gusano o una polilla, pero no un pulpo: tiene la misma protección de la que gozan los mamíferos, las aves y los reptiles. ¿Debería ampliarse o restringirse esta ley? ¿Los legisladores la establecieron adecuadamente, la primera vez? Si queremos una respuesta defendible a esta cuestión, necesitamos identificar (y luego delimitar) nuestros presentimientos. No debemos dejar que nuestras intuiciones morales distorsionen nuestra investigación empírica desde su mismo inicio.

No debemos subestimar los hitos conseguidos por agentes ciegos. El castillo de las termitas, el ovicidio del polluelo de cuco y muchas otras maravillas conductuales se logran solamente con el tipo de comprensión conductual que equivale a una experiencia práctica, sin articular y sin considerar. Cuando nosotros, observadores, explicadores y predictores humanos nos enfrentamos a esta excelencia bien diseñada, nos ponemos a descifrar las

razones por las que plantas y animales hacen lo que hacen, les aplicamos ingeniería inversa con la ayuda del enfoque intencional. Y, tal como hemos visto, cuando lo hacemos, es habitual y natural atribuir más comprensión a un organismo de la que realmente tiene, sobre la base razonable de que su comportamiento es claramente inteligente y ¿de quién va a ser esta inteligencia, si no del organismo? Resulta irónico que, si fuéramos creacionistas, podríamos atribuir con comodidad toda esta comprensión a Dios y no nos sentiríamos obligados a otorgarla a los organismos; podrían ser marionetas de Dios. Fue el descubrimiento de Darwin y la revelación del proceso ciego de selección natural, con su poder de generar motivaciones flotantes, lo que liberó nuestra imaginación para continuar aplicando la ingeniería inversa a todas las maravillas de la naturaleza sin que nos sintiéramos obligados a identificar una mente que albergara las razones que íbamos descubriendo.

¿CÓMO CONSIGUEN LOS CEREBROS HUMANOS UNA COMPRENSIÓN «GLOBAL» A PARTIR DE COMPETENCIAS «LOCALES»?

Los hombres recibieron el lenguaje de modo que así pudieran ocultar sus pensamientos.

C

HARLES-MAURICE DE TALLEYRAND

El lenguaje, como la conciencia, solo surge de la necesidad de relación con los demás.

K

ARL MARX

En general, la conciencia solo se ha desarrollado bajo la presión de la necesidad de comunicación.

F

RIEDRICH NIETZSCHE

No hay ningún general Leslie Groves para organizar y dirigir a las termitas en una colonia de termitas, ni hay tampoco ningún general Leslie Groves para organizar y dirigir a las aún más ciegas neuronas en un cerebro humano. ¿Cómo puede ser que la comprensión humana esté formada por actividades de neuronas sin comprensión? Además de todas las motivaciones flotantes que explican nuestras diversas estructuras, hábitos y demás características, hay las razones ancladas que nos representamos a nosotros mismos y a los otros. Para nosotros, estas razones son cosas, moradores de nuestra imagen manifiesta junto a los árboles, las nubes, las puertas, las tazas, las voces, las palabras y las promesas que conforman nuestra ontología. Con estas razones podemos hacer cosas (desafiarlas, reconsiderarlas, abandonarlas, validarlas, desaprobarlas), comportamientos, a menudo ocultos, que no estarían en

nuestro repertorio si no hubiéramos descargado todas las aplicaciones del lenguaje en nuestros cerebros. En resumen, podemos pensar acerca de estas razones, buenas y malas, lo que les permite influir en nuestros comportamientos públicos de maneras desconocidas para otros organismos.

La maniobra de distracción del chorlitejo silbador y su danza del ala rota da al zorro una razón para modificar su curso y acercarse al pájaro, pero no para confiar en él. El chorlitejo puede modular su actuación para captar la atención del zorro, pero el control de esta modulación no exige que el ave tenga más que una rudimentaria «apreciación» del estado mental del zorro. Por su parte, el zorro no necesita tener más comprensión de por qué se enfrasca en esta misión en lugar de proseguir su exploración de la zona. Del mismo modo, nosotros podemos efectuar muchas acciones ingeniosas y justificables retrospectivamente con solo una vaga idea de lo que estamos haciendo, una idea que a menudo se perfila a posteriori, autoatribuyéndonos razones. Y es este último paso el que es exclusivamente nuestro.

Nuestros hábitos de autojustificación (autoapreciación, autodisculpa, autoconsuelo, autoglorificación, etc.) son maneras de comportarse (maneras de pensar) que adquirimos mientras llenamos nuestras cabezas con memes culturales, entre los que se incluyen los importantes hábitos de autoacusación y autocrítica. Así aprendemos a planificar por avanzado, a usar la práctica de aventurar razones y de criticar razones para resolver por adelantado algunos de los problemas de la vida, hablando sobre ellos con otras personas y con nosotros mismos. Y no solo se trata de hablar sobre ellos, también de imaginarlos, de probar variantes en nuestra mente y de buscar fallos. No somos criaturas meramente popperianas, somos criaturas gregorianas (véase capítulo 5) que usamos recursos de pensamiento para diseñar nuestros actos futuros. No hay ningún otro animal que lo haga.

Nuestra capacidad de efectuar este tipo de pensamiento no se debe a ninguna estructura cerebral especializada que no se halle en otros animales. En otras palabras, no hay ningún «núcleo explicador». Nuestro pensamiento es posible gracias a la instalación de una máquina virtual formada por máquinas virtuales formadas por máquinas virtuales. El objetivo de perfilar y explicar este conjunto de competencias solamente mediante procesos neurocientíficos de abajo arriba (sin ayuda de la

neurociencia cognitiva) es tan distante como el de perfilar y explicar el conjunto de aplicaciones de un teléfono móvil mediante un desciframiento de abajo arriba del diseño de los circuitos de hardware y de las cadenas de bits en la memoria sin echar un vistazo a la interfaz de usuario. La interfaz de usuario de una aplicación existe para que la competencia sea accesible a los usuarios (las personas) que no saben, ni falta que les hace, los intricados detalles de su funcionamiento. Las ilusiones de usuario de todas las aplicaciones almacenadas en nuestros cerebros existen por la misma razón: hacen que nuestras competencias sean más o menos accesibles a los usuarios (otras personas) que no saben, ni falta que les hace, sus intricados detalles. Y luego las usamos nosotros mismos, aproximadamente en las mismas condiciones, como huéspedes en nuestros cerebros.

Puede que haya alguna otra ruta evolutiva (genética, no cultural) hasta una ilusión de usuario similar en otros animales, pero no he sido capaz de concebir ninguna con un mínimo detalle, y según los argumentos del etólogo y roboticista David McFarland (1989), «la comunicación es el único comportamiento que requiere que un organismo supervise su propio sistema de control». Los organismos pueden controlarse a sí mismos de forma muy eficaz mediante un conjunto de controladores de tareas en competición pero «miopes», cada uno activado por una condición (hambre u otra necesidad, sensación de la oportunidad, prioridades incorporadas, etc.). Cuando la condición que activa a un controlador pesa más que la condición del controlador actualmente activo, lo interrumpe y se pone al mando temporalmente (el «modelo del pandemónium» de Oliver Selfridge [1959] es el antecesor de muchos modelos posteriores). Los objetivos se representan solo tácitamente, en los bucles de retroalimentación que dirigen a cada controlador, pero sin ninguna representación global o de alto nivel. La evolución tenderá a optimizar la dinámica de interrupciones de estos módulos sin que nadie se dé cuenta; es decir, no es necesario que haya nadie para darse por enterado.

McFarland afirma que la comunicación es la innovación conductual que lo cambia todo. La comunicación requiere una especie de centro de coordinación para proteger al organismo de revelar demasiado acerca de su estado actual a los organismos competidores. Tal como mostraron Dawkins y Krebs (1978), para entender la evolución de la comunicación necesitamos verla como basada en la manipulación más que como un

comportamiento puramente cooperativo. Un organismo que no pueda poner «cara de póquer», que comunique indisimuladamente su estado a todos los receptores, es una presa fácil y no tardará en extinguirse (von Neumann y Morgenstern 1944). Lo que debe evolucionar para evitar esta exposición es una especie de barrera de control de la comunicación privada que cree oportunidades para un engaño guiado (y, casualmente, oportunidades para el autoengaño [Trivers 1985]) creando, por primera vez en la evolución de los sistemas nerviosos, representaciones de su estado actual explícitas y más globalmente accesibles, representaciones que se pueden separar de las tareas que representan, de modo que los comportamientos engañosos se puedan formular y controlar sin interferir con el control de otros comportamientos.

Es importante darse cuenta de que, por comunicación, McFarland no entiende una comunicación específicamente lingüística (que es solo nuestra), sino comunicación estratégica, que facilita el espacio crucial entre los objetivos propios reales y las intenciones y objetivos que uno pretende comunicar a una cierta audiencia. Sin duda, muchas especies están genéticamente pertrechadas con comportamientos comunicativos comparativamente simples (Hauser 1996), como el salto de rebote, las llamadas de aviso y las marcas y la defensa territoriales. El engaño estereotípico, como el faroleo en un encuentro agresivo, es habitual, pero un talento más productivo y versátil para el engaño necesita el espacio de trabajo privado de McFarland. Durante más de un siglo, los filósofos han subrayado la «privacidad» de nuestros pensamientos, pero pocas veces se han parado a preguntar por qué se trata de un rasgo de diseño tan bueno (es una ceguera laboral de muchos filósofos: consideran que la imagen manifiesta viene dada y no se molestan en preguntar para qué nos podría haber sido dada).

¿CÓMO SE NOS HIZO MANIFIESTA NUESTRA IMAGEN MANIFIESTA?

Aquí nos encontramos con otra inversión extraña: esta práctica de compartir información en acciones comunicativas con los demás, ofreciendo y pidiendo razones, es lo que crea nuestras ilusiones de

usuario personales. Todos los organismos, desde las células a los elefantes, tienen un rudimentario «sentido de uno mismo». La ameba esquiva con habilidad lo malo y deja entrar a lo bueno, protegiendo sus límites vitales. La langosta sabe lo suficiente como para no arrancarse y comerse sus propias extremidades. Las motivaciones flotantes del comportamiento de todos los organismos están organizadas alrededor de la protección propia. En nuestro caso, los comportamientos incluyen un conjunto de comportamientos de pensamiento ocultos que recogemos en el curso de la enculturación, un proceso que exige mucha interacción pública con los conespecíficos. La práctica lleva a la perfección, y el perfeccionamiento y ampliación de estos talentos depende, en grado extremo, de la mutua accesibilidad. El jugueteo de los cachorros y de las crías de oso afina sus capacidades de percibir y anticipar los movimientos de los demás y de percibir y modular sus propias acciones y reacciones, una buena preparación para las actividades más serias de la edad adulta. Nosotros, humanos, necesitamos desarrollar una relación similar con los demás a medida que aprendemos a comunicarnos, lo que exige que nos percibamos a nosotros mismos ejecutando estos comportamientos. Es esto lo que nos da un sentido de uno mismo menos rudimentario, más «mísmico». Necesitamos hacer un seguimiento no solo de qué extremidades son nuestras y de lo que estamos haciendo con ellas, sino también de qué pensamientos son nuestros y de si deberíamos compartirlos con los demás. Podemos dar un giro casi paradójico a esta curiosa idea: ser tú es como ser algo porque se te ha posibilitado decirnos (o abstenerte de decirnos) qué es ser como tú, qué se siente al ser como tú.

Cuando evolucionamos en un nosotros, una comunidad comunicadora de organismos que podían comparar impresiones, nos convertimos en beneficiarios de un sistema de ilusiones de usuario que reproducían versiones de nuestros procesos cognitivos (por lo demás tan imperceptibles como nuestros procesos metabólicos), accesibles a nosotros para propósitos de comunicación. McFarland no es ni mucho menos el primero en expresar la idea de que explicarnos a los otros es la actividad innovadora que genera la I+D que crea la arquitectura de la conciencia humana, tal como muestran las citas al principio de este apartado. La idea pretende ofrecer una base para la largamente esperada explicación de la evolución de la conciencia humana. Aunque esté equivocada, por lo menos proporciona un modelo de lo que debería lograr una descripción satisfactoria. Varios pensadores se han dirigido a

ideas relacionadas y aliadas: entre ellas tenemos los «símbolos activos» de Douglas Hofstadter (1979, 1982b, 1985, 2007), y tres libros de 2013, del psicólogo Matthew Lieberman, el neurocientífico Michael Graziano y el filósofo de la ciencia cognitiva Radu Bogdan.

La evolución de los memes ofrece las condiciones para la evolución de una interfaz de usuario que haga «visibles» los memes al «uno mismo», al «yo» que (quien) se comunica con los demás, el yo como el centro de gravedad narrativa (Dennet 1991), autor de palabras y de hechos. Si se necesita atención conjunta en un tema compartido (véase la discusión de Tomasello en el capítulo 12), tiene que haber cosas, ofrecimientos, a las que tanto la primera como la segunda persona puedan prestar atención, y es precisamente esto lo que hace que nuestra imagen manifiesta sea manifiesta para nosotros. Si no tuviéramos que ser capaces de hablar con los demás sobre nuestros pensamientos y proyectos actuales, sobre nuestros recuerdos de cómo eran las cosas, etc., nuestros cerebros no gastarían tanto tiempo, energía y materia gris en un boletín revisado de nuestras actividades actuales, que es precisamente lo que es nuestro flujo de conciencia. El uno mismo, el yo que tiene un acceso limitado a lo que sucede en su cerebro está bien diseñado para albergar nuevos memes, propagar viejos memes y comparar aspectos con los demás. ¿Y qué es este yo? No se trata de una parte especializada de los circuitos neurales, sino más bien el usuario final de un sistema operativo. Tal como lo expresa Daniel Wegner, en su revolucionario libro The Illusion of Conscious Will (2002), «nos es totalmente imposible saber (ni hablemos ya de hacer un seguimiento) la enorme cantidad de influencias mecánicas en nuestro comportamiento porque habitamos en una máquina extraordinariamente compleja» (p. 27). ¿No resulta remarcable la facilidad con la que seguimos a Wegner en esta visión aparentemente dualista de nosotros mismos como ocupantes inequívocos de nuestros cuerpos? Estas máquinas en las que «habitamos» simplifican las cosas en beneficio nuestro: «La experiencia de la voluntad, pues, es la manera en que nuestras mentes nos representan su funcionamiento, no su funcionamiento real» (p. 96).

Así, resulta curioso que nuestro punto de vista en primera persona acerca de nuestras mentes no sea muy diferente de nuestro punto de vista en segunda persona acerca de las mentes de los otros: no vemos, ni oímos ni sentimos la compleja maquinaria neural que se agita en nuestros cerebros, sino que tenemos que conformarnos con una versión

interpretada y masticada, una ilusión de usuario que nos resulta tan familiar que la tomamos no solo por la realidad, sino por la realidad más indudable e íntimamente conocida de todas. No se trata de una idea nueva, pero parece ser que se redescubre una vez tras otra. El gran neurólogo John Hughlings Jackson dijo en una ocasión que «hablamos no solamente para decir a los demás lo que pensamos, sino también para decirnos a nosotros mismos lo que pensamos» (1915). Yo y muchos otros hemos citado incorrectamente al novelista y crítico E. M. Forster diciendo «¿Cómo voy a saber lo que pienso hasta ver lo que digo?» Aunque Forster tiene una versión de esta cita en su libro de reseñas Aspectos de la novela (1927), su intención es sarcástica y se refiere a una anécdota anterior en la que se basa. La mutación viral del meme de Forster se ha propagado mucho, según R. J. Heeks (2013), quien demuestra que la cita, en el contexto pertinente, intenta denigrar el método de escritura de André Gide:

Hay otro distinguido crítico que coincide con Gide: la ancianita de la anécdota a quien sus sobrinas acusan de ser ilógica. No había modo de hacerle comprender lo que era la lógica, y cuando captó su verdadero significado, su reacción no fue tanto de irritación como de desprecio.

—¡La lógica! ¡Válgame Dios! ¡Qué majadería! —exclamó—. ¿Cómo voy a saber lo que pienso hasta ver lo que digo?

Sus sobrinas, educadas jovencitas, pensaron que estaba passée; pero estaba más a la moda que ellas. (Forster 1927, p. 46)

Me congratula poder corregir el error (hasta donde puedo, porque no he podido hallar pistas sobre la anécdota de la señora y sus sobrinas), pero diría que Forster rozó una posibilidad importante, si bien antiintuitiva, sin darse cuenta. Nuestro acceso a nuestro propio pensamiento y, especialmente, a la causalidad y la dinámica de las partes subpersonales no es mejor que nuestro acceso a nuestros procesos digestivos; tenemos que confiar en el canal, angosto y muy revisado, que responde a nuestra incesante curiosidad con comunicados fáciles de manejar, solo una pizca más cerca de mi yo real que el acceso a mi yo real que tienen mis amigos y familiares. Una vez más, la conciencia no es solo hablar con uno

mismo, sino que incluye también todas las variedades de autoestimulación y reflexión que hemos adquirido y perfeccionado durante la vida. No se trata simplemente de cosas que pasan en nuestros cerebros, son comportamientos en los que nos enfrascamos (Humphrey 2000, 2006, 2011), algunos «instintivamente» (gracias a la evolución genética) y el resto adquiridos (gracias a la evolución y la transmisión cultural y a la autoexploración individual).

¿POR QUÉ EXPERIMENTAMOS LAS COSAS DEL MODO EN QUE LO HACEMOS?

Si, tal como dice Wegner, «nuestras mentes nos representan [la cursiva es mía] su funcionamiento» y si, tal como acabo de comentar, nuestra conciencia individual es como la ilusión de usuario en una pantalla de una computadora, ¿no implica ello, en resumidas cuentas, la existencia de un teatro cartesiano en que se efectúa esta representación, en el que se desarrolla el espectáculo, como la obra que ves en la pantalla de la computadora? No, pero explicar lo que hay que poner en lugar del teatro cartesiano nos obligará a un cierto esfuerzo de imaginación.

Podemos hacer una lista de las propiedades de los elementos del escritorio de la computadora: «archivos» rectangulares de color azul; un cursor negro en forma de flecha; una palabra resaltada en amarillo y escrita en Times New Roman 12; etc. ¿Cuáles son las propiedades correspondientes de estos elementos, o casos, internos, privados e identificables en nuestros cerebros? No lo sabemos… todavía. En el capítulo 9 reflexionamos sobre la manera en que los significados puros, sin palabras asignadas, pueden ocupar nuestra atención en la conciencia, especialmente el fenómeno «en la punta de la lengua». Son casos genuinos, casos de memes o de tipos de sensaciones con los que nacemos o de otros ofrecimientos rememorados que pueden reconocerse y re-identificarse aunque (aún) no tengan nombres. Cerrad los ojos e imaginad una A mayúscula de color azul. ¿Ya? Habéis creado un caso en vuestro cerebro, pero podemos estar seguros de que no es azul, no más que los casos de una «o» que se producen en un archivo de un procesador de textos son redondos. Las «casificaciones» se dan en la

actividad de los circuitos neuronales y tienen un papel importante en centrar la atención, en despertar casos relacionados y en modular muchas actividades cognitivas. Contribuyen a dirigir acciones básicas como las sacudidas oculares y a iniciar acciones de alto nivel como la activación de docenas de aplicaciones (memes) que, como siempre, están empeñadas en colocar nuevos casos de sí mismas (descendencia) en el terreno de juego. Fijaos:

cebr ryda

Es posible que la experiencia visual que acabo de proporcionar haya despertado en vuestras mentes las palabras cebra y rayada y quizá (¿os habéis dado cuenta?) estos casos tenían un molde «auditivo» concreto, con las tres aes de las palabras destacándose frente al estímulo visual casi impronunciable que las ha despertado. Estas palabras han poblado vuestros espacios de trabajo neurales con representaciones de rayas blancas y negras que, por supuesto, no son ellas mismas blancas ni negras (¿erais conscientes de las rayas en vuestra imaginación visual? quizá no, porque la activación no era lo bastante fuerte en vuestro caso concreto, pero podéis estar seguros de que los casos subpersonales, y subconscientes, se han activado, ya que «prepararían» vuestras respuestas a otras preguntas en un entorno experimental).

En esta actividad neural, subpersonal, es donde se producen las interacciones causales de verdad, que proporcionan los poderes cogniti-

vos, pero cada uno de nosotros solamente tiene acceso a los resultados. Mediante la introspección no puedes descubrir cómo cebr se convirtió en cebra

y, luego, en una imagen mental de una cebra, con sus rayas. Cuando intentamos explicar a los demás lo que está sucediendo con esta experiencia, nos vemos impelidos a usar imágenes metafóricas, simplemente porque no disponemos de un conocimiento más profundo, más verdadero y más exacto de lo que está pasando en el interior de cada uno. Acomodamos nuestra ignorancia con un modelo falso, pero

tentador: reproducimos, con elucubraciones y disculpas, nuestro modelo cotidiano de cómo sabemos lo que pasa en nuestro exterior.

El proceso es como sigue. Empecemos recordando algo familiar y bien entendido: enviamos a un periodista a observar alguna parte del mundo externo (una casa cercana, pongamos por caso) y que nos informe por teléfono. Suena nuestro teléfono y cuando contestamos, nos dice que la fachada tiene cuatro ventanas; le preguntamos cómo lo sabe y nos espeta «¡Porque las estoy viendo en este mismo momento, a plena luz del día!» Normalmente no pensamos en preguntar cómo el hecho de que las vea, claras como el agua, explica cómo sabe este hecho. Ver es creer, o algo parecido. De una forma tácita, asumimos que el recorrido desconocido desde sus ojos que ven hasta sus labios que hablan es seguro, igual que la actividad necesaria en las torres de telecomunicaciones entre su teléfono y el nuestro. En una situación así nunca nos preguntamos cómo funcionan los teléfonos; lo damos por sentado. Del mismo modo, tampoco nos devanamos los sesos, perplejos, preguntándonos cómo el periodista puede abrir los ojos y luego responder preguntas con confianza acerca de lo que tiene ante ellos, a plena luz del día, porque todos podemos hacerlo (los que no somos ciegos, claro está). ¿Cómo funciona todo esto? No lo sabemos y, en general, no nos lo preguntamos.

Si nos ponemos curiosos y le preguntamos que nos describa, no el mundo externo, sino su experiencia subjetiva del mundo externo, su mundo interior, lo ponemos en un aprieto al pedirle que haga una cosa bastante poco natural; por ende, los resultados suelen ser decepcionantes (a no ser que sea un «introspector avezado» de una escuela u otra): «No sé… Miro y veo la casa. O sea, creo que veo una casa; hay una cosa en forma de casa que parece que está a unos cincuenta metros, y tiene cuatro cosas que parecen ventanas… si cierro los ojos y los vuelvo a abrir, sigue ahí…».

La relativa accesibilidad y familiaridad de la parte externa del proceso de decir a los demás lo que podemos ver (sabemos que tenemos que abrir los ojos, centrar la mirada y estar atentos, y tiene que haber luz) nos oculta el absoluto vacío (desde la perspectiva de la introspección o la mera autoexploración) del resto del proceso. No tenemos un acceso más privilegiado a esta parte del proceso del que tenemos a los procesos complejos que mantienen la conexión entre el teléfono del periodista y el nuestro.

—¿Cómo sabes que hay un árbol al lado de la casa?

—Bueno, ahí está, y puedo ver que tiene todo el aspecto de un árbol.

—¿Cómo sabes que tiene el aspecto de un árbol?

—¡Lo sé, y ya está!

—¿Comparas su aspecto con muchas otras cosas del mundo antes de llegar a la idea de que es un árbol?

—No, conscientemente no.

—¿Lleva una etiqueta «árbol»?

—No, no necesito ver ninguna etiqueta; además, si hubiera una etiqueta tendría que leerla y saber que es una etiqueta para la cosa sobre la que está. Simplemente, sé que es un árbol.

Ahora imaginemos que al separar los dedos de los pies podemos llegar a tener una idea increíblemente precisa de lo que está pasando ahora mismo en Chicago. E imaginemos también que no tenemos curiosidad por saber cómo es esto posible.

—¿Cómo lo haces?

—Ni idea, pero funciona, ¿verdad? Si junto los dedos de los pies, no funciona, pero si los abro otra vez, cualquier curiosidad que tenga por lo que está pasando ahora mismo en Chicago, se me revela al instante. Lo sé, así, sin más.

—¿A qué se parece?

—Bueno, es como ver y oír, como si estuviera viendo una emisión de televisión, pero tampoco es exactamente así. Simplemente, mi curiosidad sobre Chicago se satisface sin ningún esfuerzo por mi parte.

La explicación tiene que acabar en algún punto, y a nivel personal, acaba aquí, con las capacidades brutas formuladas en el familiar lenguaje mentalista de saber, ver, darse cuenta e identificar. El problema con el punto de vista en primera persona es que está anclado en la imagen manifiesta, no en la imagen científica, y no puede servirse de los recursos de la imagen científica. La suposición habitual es que cuando cuestionamos al periodista, la respuesta «lo sé porque puedo verlo» es una respuesta lo bastante completa, pero cuando traspasamos esta misma suposición al caso en que una persona informa basándose en la imaginería mental o la memoria (o la inventada clarividencia sobre Chicago con los dedos de los pies), por ejemplo, creamos un artefacto. Lo que nuestras preguntas crean, o provocan, directamente son respuestas, como en los diálogos anteriores; lo que crean indirectamente son ideas basadas en esas respuestas. Podemos preguntarnos cuál es nuestra experiencia subjetiva y ver lo que queremos decir; luego podemos decidir respaldar nuestras propias afirmaciones, creerlas y seguir sus implicaciones. Esto lo podemos hacer sin más que hablar en voz alta con nosotros mismos, o en silencio, o simplemente «pensando» sobre lo que estamos experimentando en el momento.

Este es el grado de acceso a nuestras propias experiencias, que no es muy diferente del acceso que otra persona puede tener a estas mismas experiencias, nuestras experiencias, si decidimos hacerlas públicas con una explicación. Nuestras convicciones son fiables, sin duda, pero no infalibles. Otra persona podría ayudarnos a probarlas y quizá a ajustarlas para experiencias futuras. Esta es la forma de estudiar la conciencia científicamente, y la he bautizado con un nombre poco agraciado pero preciso: heterofenomenología, la fenomenología de la experiencia de otra persona, a diferencia de la autofenomenología, la fenomenología de la experiencia propia. Hay una larga tradición que afirma que, de algún modo, la autofenomenología es una manera más íntima, más auténtica y más directa de llegar a los objetos de la experiencia, que adoptar el punto de vista en primera persona es la jugada clave en cualquier estudio prometedor de la conciencia; pero se trata de un espejismo. La heterofenomenología es más precisa, más fiable y menos vulnerable a las ilusiones que la autofenomenología, una vez se tiene el control de las posibles mentiras y otras formas de falta de cooperación en la investigación; además, se puede obtener un mejor catálogo de las

experiencias propias sometiéndose uno mismo a todas las situaciones experimentales en que se estudia la conciencia. Pueden mostrarse ante uno mismo rasgos de la experiencia propia de los cuales no se tenía ningún conocimiento, tanto carencias y debilidades inesperadas como capacidades sorprendentes que uno mismo no sabía que tenía.

Colaborar con otros investigadores en el estudio de la conciencia propia (es decir, adoptar el punto de vista en segunda persona) es la forma de considerar el fenómeno de la conciencia de la manera más seria posible. Resistirse a esto insistiendo que yo sé más acerca de mi propia conciencia simplemente porque es la mía, es caer en el dogma; al blindar mi valiosa experiencia a cualquier tipo de sondeo, perpetúo mitos que ya no tienen utilidad.

Pongámonos en otra situación: pedimos a una persona que nos diga cuántas ventanas había en la habitación en que dormía de pequeño; cierra los ojos un instante y responde «dos». Le preguntamos «¿Cómo lo sabes?» y nos responde «Bueno, he "mirado"… y las he "visto"». Está claro que no ha mirado, en un sentido literal; sus ojos estaban cerrados (o tenía la mirada perdida a media distancia). La parte de los ojos del proceso de mirar no estaba activada, pero buena parte del resto de la visión sí (la parte que no solemos cuestionar). Es un poco como ver y no ver, pero exactamente cómo funciona no resulta accesible a una exploración de psicología popular, a la introspección o a la automanipulación. Cuando nos enfrentamos a este familiar vacío, tenemos una tentación casi irresistible de postular un universo sustituto (una imagen mental) para reemplazar la parte del mundo real que observa el periodista. Y podemos estar seguros de la existencia de este mundo sustituto, al menos en un sentido forzado: tiene que haber algo (algo en la actividad neuronal) que mantenga de forma fiable y consistente una gran cantidad de información sobre el tema en cuestión, pues podemos confirmar fácilmente el hecho de que de «él» se puede obtener información casi con tanta fiabilidad como con la observación de alguna cosa del mundo real. La «imagen rememorada» de la casa tiene una cierta riqueza y precisión que puede comprobarse y se pueden evaluar sus límites. Estos límites nos dan pistas importantes sobre cómo se encarna realmente la información en el cerebro; pero tampoco debemos apresurarnos a sacar la conclusión de que está encarnada, como parece, en una imagen que se puede consultar.[2]

Desde esta perspectiva, nuestra absoluta incapacidad para decir lo que estamos haciendo cuando hacemos lo que denominamos encuadrar imágenes mentales no resulta sorprendente. Aparte de las cosas accesorias que estamos haciendo con nuestros ojos, somos igual de incapaces de decir lo que estamos haciendo al ver el mundo externo. Simplemente miramos y aprendemos, y esto es todo lo que sabemos. Consideremos los procesos subpersonales de la visión normal y fijémonos en que en algún punto tienen que poder explicar todas las cosas que podemos hacer gracias al hecho de tener los ojos abiertos: coger moras, chutar pelotas, identificar paisajes, navegar por terreno desconocido o leer, por ejemplo. Además, gracias a estos procesos, nuestros estados corticales internos bastan para dirigir nuestro subsistema del habla hacia el encuadre de actos de habla descriptivos. Estamos haciendo progresos constantes en esta historia subpersonal, aunque aún quedan muchas partes desconcertantes. Podemos tener bastante confianza en que habrá una historia subpersonal que empezará en los ojos y recorrerá todo el camino hasta las descripciones orales (entre otras cosas) y en tal historia no habrá un segundo proceso de presentación con un ego (un yo, un jefe, un testigo interno) observando una pantalla interna y luego elaborando una descripción. Nunca me cansaré de decir que todas las tareas efectuadas por el homúnculo imaginario del teatro cartesiano tienen que dividirse y distribuirse (en el espacio y en el tiempo) por departamentos más pequeños del cerebro.

En tal caso, pues, intentemos dividir el yo en algunas partes plausibles. ¿Qué reducciones, qué cortes del periodista observador podrían resolver el problema? Quizá «agentes» llenos de convicciones pero sin ninguna idea de cómo las obtuvieron; más bien como oráculos, acaso, plagados de valoraciones pero sin nada que nos pudieran decir (o que se pudieran decir a sí mismos) acerca de cómo llegaron a ese estado de convicción. Ray Jackendoff se ocupó de este problema hace unos años (1996) y presentó una posibilidad interesante, desarrollada recientemente por Bryce Huebner y yo mismo, que introduce el concepto de emisión subpersonal o blurt subpersonal (2009):

La percepción clave es que un módulo «convierte, ciega y obsesivamente, pensamientos en forma lingüística y viceversa» (Jackendoff 1996). Esquemáticamente, un pensamiento conceptualizado

activa la producción de una representación lingüística que hace una aproximación al contenido de ese pensamiento y genera un blurt reflexivo. Estos blurts lingüísticos son actos de protohabla, que se emiten subpersonalmente, aún no proceden de la persona ni son emitidos por ella, y o bien se envían a sistemas de transmisión exógenos (donde se convierten en materia prima para actos personales de habla) o bien se transmiten de modo endógeno a los sistemas de comprensión del lenguaje, que los envían directamente al sistema de lectura de la mente. Aquí, los blurts se comprueban para ver si deben expresarse abiertamente, a medida que el sistema de lectura de la mente accede al contenido del blurt y, reflexivamente, genera una creencia que es una aproximación al contenido de ese blurt. A continuación, se reclutan sistemas dedicados a la fijación de las creencias, las creencias se actualizan y el blurt es aceptado o rechazado; luego, se repite el proceso. (Huebner y Dennett 2009, p. 149)

Toda esta descripción es de trazo muy grueso y necesita más detalles, pero una idea similar de creación de «narrativa» ha sido elaborada recientemente por Gustav Markkula (2015), que argumenta con convicción que la actividad humana de preguntarnos y explicarnos a nosotros mismos qué es ser como nosotros, crea artefactos de la imaginación que consideramos que son los qualia tan apreciados por los filósofos de la conciencia que anhelan reinstaurar el dualismo como una teoría seria de la mente.

LA EXTRAÑA INVERSIÓN DEL RAZONAMIENTO DE HUME

Todavía queda una objeción: ¿por qué tiene que ser como algo ver, oír y oler? ¿Por qué parece haber un teatro interior con un espectáculo multimedia en marcha siempre que estamos despiertos? Incluso concediendo que tiene que haber una historia subpersonal, en la imagen científica, que puede explicar satisfactoriamente todos los comportamientos y respuestas emocionales, las decisiones y las

expresiones verbales que yo genero, ¡debe dejarme fuera de la historia! Y colocarme, a mí y mis qualia, otra vez en el mundo, sin dejarlos fuera, es una tarea pendiente. La mejor respuesta que conozco a este problema es lo que denomino la extraña inversión del razonamiento de Hume, porque fue él quien formuló una descripción profética de un caso (nuestra experiencia de la causalidad) mucho antes de que Darwin y Turing descubrieran sus respectivas inversiones. Las complicaciones y controversias acerca de la teoría de la causalidad de Hume son legión, y parte de su descripción, si bien ha sido muy influyente durante siglos, actualmente está casi abandonada; sin embargo, hay una idea central que destaca y que proporciona una importante apreciación sobre la relación entre las imágenes manifiesta y científica y sobre la naturaleza de nuestra experiencia consciente en general, no solo de nuestra experiencia de causalidad.

Nos parece ver, oír y notar la causalidad cada día, afirma Hume, como cuando vemos una piedra que rompe una ventana u oímos un timbre cuando lo pulsamos, pero todo lo que experimentamos directamente, prosigue Hume, es una secuencia: A seguido por B, no A que causa B. Si Hume estuviera equivocado, los dibujos animados serían imposibles: al representar a Bugs Bunny mordisqueando una zanahoria, los animadores no solo tendrían que añadir una pista de sonido, con un fuerte ñam sincronizado con el mordisco, sino también una especie de «pista de causalidad» que, de algún modo, nos mostrara directamente que cuando los dientes de Bugs Bunny muerden, causan y no solo preceden a, la desaparición de media zanahoria y el soni-

do de mordisco que oímos. Por supuesto, todo esto no es necesario; la sucesión de un fotograma y el siguiente es todo lo que se necesita para crear la impresión de causalidad. Pero tal como señala Hume, la impresión de causalidad que experimentamos procede del interior, no del exterior; es un efecto de un hábito de expectativas que se nos ha inculcado a lo largo de horas y horas de vigilia (Hume insistía en que estos hábitos de expectativas eran aprendidos, adquiridos durante la infancia, pero la investigación actual sugiere que nacemos con una especie de sentido causal automático, como un reflejo, listo para «ver» causación siempre que nuestros sentidos se enfrenten a la secuencia adecuada de estímulos). Cuando vemos A, estamos programados para esperar B; así, cuando sucede B (he aquí la jugada maestra de Hume), atribuimos incorrectamente nuestra reacción perceptual a alguna causa

externa que estamos experimentando más o menos directamente (creemos que vemos los dientes dibujados de Bugs Bunny causar el corte de la zanahoria). De hecho, sucumbimos a una ilusión de usuario benigna, interpretando equivocadamente nuestra expectativa satisfecha de un B subsiguiente como si procediera del mundo exterior. Siguiendo a Hume, se trata de un caso especial de «gran propensión de la mente a extenderse en los objetos externos» (1739, I:xiv). La «transición habitual» en nuestras mentes es el origen de nuestro sentido de la causalidad, una calidad de las «percepciones, no de los objetos» y, como señala, «la idea opuesta está tan afianzada en la mente» que es difícil de extraer. Sobrevive hasta el día de hoy en la habitual suposición indiscutida de que todas las representaciones perceptuales deben proceder del exterior.

A continuación, presento otras convicciones populares que necesitan someterse a la extraña inversión de Hume: la dulzura es una propiedad «intrínseca» del azúcar y de la miel, que provoca que nos guste; la atracción sexual observada es lo que causa nuestro deseo; la gracia de un chiste es lo que nos mueve a la risa (Hurley, Dennett, Adams 2011). Simplificando en exceso, en estos casos las causas y los efectos de la imagen manifiesta se han trocado en la imagen científica. No se puede hallar ninguna dulzura intrínseca estudiando la estructura molecular de la glucosa; hay que ir a por los detalles en los cerebros de los «buscadores de dulzura». Cómo responden nuestros cerebros es lo que nos causa (en la imagen manifiesta) que proyectemos una propiedad ilusoria en el mundo (manifiesto). Hay propiedades químicas y estructurales de la glucosa (imitada en la sacarina y otros edulcorantes artificiales) que provocan la respuesta de dulzura en nuestros sistemas nerviosos, pero la «dulzura intrínseca y subjetiva de la que disfruto» no es ninguna recreación interna ni ningún modelo de esas propiedades químicas, ni tampoco es ninguna propiedad muy especial de nuestras mentes incorpóreas con las que solemos decorar las cosas perceptibles del mundo exterior. No es ninguna propiedad, solo es una ilusión benigna. Nuestros cerebros nos han engañado para que tengamos la convicción, la opinión de que parece haber una propiedad intrínsecamente genial pero por lo demás indescifrable en algunos objetos comestibles: la dulzura. La podemos identificar, recordar y soñar, pero no la podemos describir; es inenarrable e inanalizable.

Para describir este efecto no hay verbo mejor que proyectar, pero todo el mundo sabe que es una metáfora; los colores no se proyectan (al modo de un proyector de diapositivas) sobre las superficies de los objetos (incoloros), en cualquier caso, no más que la idea de causalidad se transmite hasta el punto de impacto entre dos bolas de billar. Si nos servimos, por comodidad, del término proyectar para hablar metafóricamente de la discordancia entre la imagen manifiesta y la imagen científica, ¿cuál es la historia de verdad? ¿Qué sucede realmente con la imagen científica? Mi propuesta es que una gran parte de la respuesta surge de la perspectiva de la codificación predictiva que exploramos brevemente en el capítulo 8 («¿Cómo recoge el cerebro los ofrecimientos?»).

Aquí es donde las expectativas bayesianas pueden tener un papel repetitivo: nuestra ontología (en el sentido del ascensor) hace un trabajo casi óptimo de catalogación de las cosas del mundo que son relevantes para los comportamientos que tienen que controlar nuestros cerebros. Las predicciones bayesianas jerárquicas logran este proceso, generando una sobreabundancia de ofrecimientos: esperamos que los objetos sólidos tengan una parte posterior que podremos observar si damos la vuelta a su alrededor; esperamos que las puertas se puedan abrir, esperamos que las escaleras nos permitan subir; esperamos que los vasos puedan contener líquidos, etc. Pero entre todas las cosas de nuestro Umwelt que son relevantes para nuestro bienestar, ¡estamos nosotros mismos! Deberíamos tener buenas expectativas bayesianas acerca de qué haremos a continuación, qué pensaremos a continuación y qué esperaremos a continuación. Y, efectivamente, es lo que hacemos. A continuación, propongo un ejemplo.

Pensemos en lo monos y lindos que son los bebés. Está claro que la lindura no es ninguna propiedad «intrínseca» de los bebés, aunque lo parezca. Lo que «proyectamos» en el bebé es, de hecho, nuestra variedad de disposiciones «sentidas» a acurrucar, proteger, criar, besar, arrullar… a esa ricura. No es solo que cuando se activa nuestro detector de lindura (basado en proporciones faciales, etc.) tengamos ansias de criar y proteger, sino que esperamos tener estas ansias, y toda la diversidad de expectativas no es más que la «proyección» sobre el bebé de la propiedad de la lindura. Cuando esperamos ver a un bebé en la cuna, también esperamos encontrarlo mono, es decir, esperamos esperar la sensación ansiosa de arrullarlo y todo lo demás. Cuando nuestras

expectativas se cumplen, nuestros cerebros interpretan la ausencia de señales de errores de predicción como una confirmación de que, en efecto, la cosa del mundo exterior con la que estamos interactuando realmente tiene las propiedades que esperábamos que tuviera. La lindura como propiedad supera la prueba bayesiana de ser una parte objetiva estructural del mundo en que vivimos, y con esto basta; cualquier otro proceso adicional de «proyección» sería redundante. Lo que resulta especial de propiedades como la dulzura o la lindura es que su percepción depende de particularidades del sistema nervioso que han evolucionado para aprovecharlas. Tienen un papel tendencioso o privilegiado en la modulación de nuestros sistemas de control… en resumen, nos preocupamos de ellas.

En este punto tenemos que ser muy cuidadosos de no mezclar dos afirmaciones independientes. Las propiedades de la dulzura y de la lindura dependen de características de nuestros sistemas nerviosos y, por ende, son subjetivas en este sentido restringido, pero ello no implica que la dulzura, por ejemplo, sea una propiedad intrínseca (subjetiva) de la experiencia consciente. La extraña inversión de Hume es fantástica pero incompleta: cuando habla de la «gran propensión de la mente a extenderse en los objetos externos», no debe pensarse como un punto final, sino como un puente hacia otra inversión. La imagen de Hume evoca con brillantez la visión curiosa de la mente pintando el mundo externo con los colores propios («intrínsecos»), llevados apropiadamente por los elementos internos de la mente: impresiones e ideas, en su vocabulario. Pero no existe tal pintura (por eso lo denominé en una ocasión producto de la imaginación).******** Tenemos que llevar la inversión de Hume un poco más allá y mostrar que los iconos de la ilusión de usuario de nuestras mentes, a diferencia de la ilusión de usuario de nuestras computadoras, no necesitan proyectarse en una pantalla.

UNA RAYA ROJA COMO UN OBJETO INTENCIONAL

Otro ejemplo servirá para aclarar mi posición, aunque quizá no para convencer a todo el mundo; al fin y al cabo, como dice Hume, la idea

opuesta está muy afianzada en nuestras mentes. Mirad fijamente la cruz blanca que hay en la imagen superior de la figura 14.1 durante unos diez segundos; luego, desplazad la mirada a la cruz blanca de la imagen inferior. ¿Qué veis?

—Veo una bandera de Estados Unidos, roja, blanca y azul.

—¿Ves una raya roja en la parte superior, a la derecha? [repetid el experimento]

—Sí, claro. Hay una raya roja borrosa y débil a la derecha de la zona azul con estrellas.

FIGURA 14.1. Imagen sucesiva de colores complementarios********

Ahora pensad: no hay ninguna raya roja en la imagen, en vuestras retinas ni en vuestros cerebros. De hecho, no hay ninguna raya roja en ningún sitio. Solo os parece que hay una raya roja. Vuestro cerebro «proyecta» una raya roja inexistente en el mundo (es importante que la raya imaginaria no parece estar en vuestras cabezas; parece estar en la página, como procedente de un proyector situado en vuestra frente). El fenómeno responsable de esto no es ninguna raya roja; es una representación de una raya roja en algún sistema neuronal de representación que aún no hemos ubicado con precisión y que aún no sabemos cómo descodificar, pero que sí sabemos que no es rojo ni es una raya. No sabéis qué es exactamente lo que provoca que veáis una raya roja en el mundo exterior, y por ello estáis tentados de caer en el error de atribución humeano: interpretáis erróneamente vuestro sentido (opinión, convicción, creencia, tendencia) de que estáis viendo una raya roja como si surgiera de una propiedad subjetiva (un quale, en la jerga filosófica) que es el origen de vuestro juicio, cuando en realidad es justo al revés. Es vuestra capacidad de describir «la raya roja», vuestro juicio, vuestra disposición a hacer las afirmaciones que acabáis de hacer y vuestras reacciones emocionales (si las hay) a la «raya roja» lo que constituye el origen de vuestro convencimiento de que hay una raya roja subjetiva.

Estamos ante un caso concreto de un tipo de error que se ha estudiado mucho en otros campos de la filosofía: confundir el objeto intencional de una creencia con su causa. En general, cuando los sentidos, o un timador, no te engañan, si crees en algo (en la existencia de algo a tu lado, con unos rasgos concretos, por ejemplo) es porque esa cosa con esas características ha causado que creas en ella, estimulando tus órganos sensoriales. Crees en la manzana que sostienes con la mano porque esa manzana ha causado que creas en su existencia, reflejando la luz hacia tus ojos y ejerciendo una cierta fuerza vertical hacia abajo sobre tu mano. En un caso normal como este, dejando de lado pequeñas objeciones, podemos decir que la manzana, el objeto intencional de tu creencia también es la causa (primaria, o destacada) de tu creencia. Pero hay casos anormales bien conocidos: espejismos, ilusiones ópticas, alucinaciones y postimágenes o imágenes consecutivas de colores complementarios… y engaños. Supongamos que un grupo de amigos decidimos hacer una jugarreta a nuestro amigo común Otto: ideamos un personaje ficticio, Dan Quale, y empezamos a enviar correos

electrónicos, mensajes de texto, cartas y felicitaciones de cumpleaños de Dan Quale a Otto, junto con huellas, llamadas telefónicas y encuentros cercanos (pero no demasiado) que preparamos para Otto con el elusivo (y ficticio) Quale. Otto no tardará en creer que Dan Quale es una persona real, con un registro biográfico reciente y unos datos físicos bastante detallados. Este Dan Quale es el objeto intencional de una diversidad de creencias de Otto. Estas creencias son todas acerca de Dan Quale, aunque este no existe. Hay muchas personas que existen, junto con sus correos electrónicos y sus huellas y todo lo demás, pero no Dan Quale. Las creencias de Otto sobre Dan Quale tienen un abanico de causas semiorganizadas, ninguna de las cuales es una persona llamada Dan Quale. Pero Otto no sabe todo esto; está bastante seguro de que Quale existe: le ha visto, ha hablado con él por teléfono, tiene cartas suyas, etc. Le gustaría encontrarse con Dan Quale, pero ¿con quién quiere encontrarse? No con ninguno de los bromistas, que Otto conoce perfectamente. Dan Quale no existe pero es el objeto intencional de la búsqueda de Otto. Más o menos como Ponce de León buscando la fuente de la eterna juventud. Ponce de León tenía una idea en su mente de la fuente de la eterna juventud, pero este estado mental no era el objeto de su búsqueda; ¡ya lo tenía! No estaba buscando una idea, estaba buscando una fuente. Y Otto no está buscando sus estados mentales acerca de Dan Quale, está buscando a un hombre, impulsado en tal búsqueda por sus estados mentales.

Ahora apliquemos el mismo análisis a la raya roja. Si no supierais nada de las imágenes sucesivas (postimágenes) de colores complementarios, podríais confiar de un modo algo ingenuo en la existencia real de una raya roja, visible también para los demás desde el mismo punto de vista, en el mundo «externo». Si es esto lo que creéis, el objeto intencional de vuestra creencia no existe y entre sus causas hay una imagen de una bandera a rayas verdes y una gran cantidad de sucesos neuronales en vuestra corteza visual, ninguno de los cuales es rojo ni parece ser rojo. No sois tan ingenuos y sabéis perfectamente que no existe tal raya roja externa, lo que os puede confundir y llevaros a la suposición (o, en muchos casos, a la firme convicción) de que no podéis estar equivocados: realmente hay una raya roja subjetiva en vuestra mente. ¡La estáis viendo! Bueno, digamos que la estáis viendo. Como respaldo de esta suposición teórica podéis preguntar: ¿Cómo podríamos tener una experiencia de una raya roja horizontal a no ser que algo, en algún lugar,

sea rojo y horizontal? La respuesta burda a esta pregunta retórica es: «Fácil, si no puedes concebirlo, inténtalo con más fuerza».

Este momento señala el nacimiento de los qualia, un producto de un mal proceso teórico. El objeto intencional de vuestras creencias no está en tela de juicio: creéis con todas vuestras fuerzas no que haya una raya roja ahí afuera, sino que hay una raya roja aquí dentro (algo con los qualia de una raya roja): al fin y al cabo, podéis «mirarla», «concentraros en ella», «recordarla», «disfrutarla», «compararla con otras cosas parecidas en la memoria». Se supone que los qualia son las propiedades subjetivas e internas con las que estamos más familiarizados, cuando estamos algo menos familiarizados con sus causas externas normales (rayas rojas reales en el mundo, entre todo lo demás). Al hacer esta jugada estáis planteando una causa interna que tiene las mismas propiedades que los objetos intencionales que suelen causar vuestras creencias perceptivas, excepto por el hecho de que son versiones privadas y subjetivas de las propiedades públicas y objetivas de la rojez, de la horizontalidad, etc. Pero cuando os dais cuenta de que los objetos intencionales de creencias erróneas no existen, en ningún sitio, no hay ninguna necesidad de plantear ni conjeturar ninguna misteriosa cosa interna con propiedades misteriosas. Dan Quale, el objeto intencional de las creencias de Otto sobre Dan Quale, no está hecho de ectoplasma, de ficticioplasma ni de nada. Como tampoco Papá Noel o Sherlock Holmes. Así, cuando os parezca ver una raya roja cuando no hay ninguna raya roja en el mundo que sea su origen, no es necesario que haya ninguna otra cosa (hecha de roja fantasía) que sea la «apariencia real» que os parece estar experimentando.

Entonces, ¿qué hay en su lugar? ¿Qué es lo que explica vuestra convicción de que hay una raya roja? Pues la presencia en vuestro cerebro de algo (sí, en efecto, tiene que haber algo en vuestro cerebro que sea el responsable), pero es algo que está en el medio en que se produce la actividad de impulsos nerviosos, no en algún otro medio: un estado subpersonal destacado y repleto de información, un caso de una representación de una raya roja que no es más rojo ni más raya que los casos neurales de palabras que describimos en el capítulo 9 eran fuertes o débiles, blancos o negros. Esta es la causa de vuestra creencia en la raya roja, pero no es el objeto intencional de vuestra creencia (porque no es rojo ni raya).

Pero incluso si esto es una posible explicación de todos mis estados subjetivos, ¿cómo sabemos que, de algún modo, no hay un medio de qualia en nuestras mentes, si no en nuestros cerebros? ¿Cómo sabemos que la teoría «ingenua» está equivocada? Supongamos que, en lugar de responder la pregunta retórica de mala manera, la aceptamos, de momento, y vemos adónde nos lleva. Supongamos que, en efecto, hay algún tipo de propiedad subjetiva que «explica» vuestras actuales capacidades y convicciones introspectivas. Supongamos que cuando experimentamos lo que parece ser una raya roja horizontal realmente hay, en alguna parte, un quale (sea lo que sea) rojo de forma horizontal y que este quale, de algún modo, es la causa o el origen de vuestra convicción de que estáis experimentando una raya roja horizontal, y que esta representación en algún medio desconocido está causada o activada por la confirmación (la ausencia de refutación) de todas las expectativas generadas por el funcionamiento normal de vuestro sistema visual. Para clarificar al máximo la suposición, he aquí una versión algo ampliada de la supuesta explicación del efecto de la imagen sucesiva roja:

Al fijarnos en las rayas verdes reales frente a nosotros durante unos segundos, se fatigan los circuitos neurales relevantes del sistema de colores complementarios, que generan una falsa señal (roja, en lugar de verde) que no recibe ninguna refutación mientras dura la fatiga; de este modo, en algún lugar bastante arriba en el proceso entre la retina y… ejem… el centro de la convicción filosófica se representa un quale en forma de raya roja, y es la apreciación de este quale lo que fundamenta, alimenta, informa, causa, subraya la convicción filosófica de que ahora mismo estamos disfrutando de un quale rojo en forma de raya.

Esto explica con detalle la idea que hay detrás de la pregunta retórica: necesitamos algo como esto (¿verdad?) para explicar el hecho innegable de que sin duda nos parece que aquí y ahora hay una raya roja. No solo lo decimos (del modo que lo haría un robot programado para ser un modelo de imágenes consecutivas de colores complementarios), lo creemos con todas las fuerzas.

Muy bien, ahora tenemos qualia instalados en nuestro esquemático modelo del proceso. ¿Qué más? Algo tendría que tener acceso a la

representación en ese medio (si no, los qualia representados serían inútiles, inapreciados, invisibles, como una hermosa pintura encerrada en una habitación vacía). Llamemos observador interior a lo que sea que dispone de este acceso. Ahora, ¿qué reacción suponemos que sería adecuada para este observador interior ante la representación? ¿Qué podría ser sino el juicio de que sin duda parece haber una raya roja ahí afuera, parte de lo que parece una bandera de Estados Unidos? Pero a esta conclusión ya habíamos llegado en el curso de las expectativas no refutadas. El sistema ya había identificado una raya roja en una ubicación concreta del espacio visual; esta información fue la que conformó la representación interna (del mismo modo que un mapa de bits conforma la representación de colores en la pantalla de una computadora). Postular los qualia solo sirve para duplicar la tarea cognitiva a realizar. No hay nada más que tenga que hacer la conciencia.

Esta es la importancia de preguntarse siempre lo que he denominado la Pregunta Difícil (1991): «Y ahora ¿qué?» Muchos teóricos de la conciencia se conforman con una teoría a medias. Si queréis una teoría completa de la conciencia, esta es la pregunta que hay que plantearse y que hay que responder tras haber entregado algo «a la conciencia» (sin importar a qué equivale eso que llamamos tomar conciencia). Si, en cambio, os detenéis ahí y os proclamáis victoriosos, habréis cargado al Sujeto o al Yo con la tarea de reaccionar, de hacer algo con la entrega, y habréis dejado esa tarea sin analizar. Si la respuesta que dais a la Pregunta Difícil se limita a repetir de forma inquietante la respuesta que disteis a las cuestiones «fáciles» sobre cómo funciona la parte del proceso previa a los qualia, podéis llegar a la conclusión de que estáis andando en círculos. Parad. Reconsiderad.

Empecinarse en seguir la idea de que los qualia son tanto las causas como los objetos intencionales (los objetos intencionales existentes) de las creencias introspectivas conduce a más fantasías artefactuales, la más extravagante de las cuales es la idea de que, a diferencia de nuestro conocimiento de todos los otros tipos de causalidad, nuestro conocimiento de la causalidad mental es directo e infalible: no podemos equivocarnos cuando afirmamos que nuestras creencias subjetivas sobre los elementos de nuestra experiencia consciente están causadas por esos mismos elementos. Tenemos «acceso privilegiado» a las causas o los orígenes de nuestras convicciones introspectivas. ¡No hay lugar lógico aquí para la intervención de estafadores! ¡No podríamos ser víctimas de

ninguna ilusión! Podríais ser zombis, dirigiéndoos inconscientemente a tener conciencia real con qualia reales, ¡pero yo sé que no soy un zombi! No, no lo sabéis. El único respaldo para esta convicción es la vehemencia de la misma convicción, y en cuanto os permitáis considerar la posibilidad teórica de que pueda haber zombis, tendréis que abandonar vuestra sacrosanta autoridad acerca de vuestra propia falta de zombicidad. No puedo probarlo, todavía, pero puedo alentar a los potenciales teóricos de la conciencia a identificar el abismo creado por esta jugada y reconocer que no se pueden satisfacer ambas posturas.

¿QUÉ ES LA GRAVEDAD CARTESIANA Y POR QUÉ SE MANTIENE?

René Descartes no fue el primer gran pensador que intentó ofrecer una descripción de la mente humana, pero su punto de vista, tal como lo presentó en su Discurso del método (1637) y las Meditaciones filosóficas (1641), era tan brillante y convincente que ha influido profundamente en todas las ideas posteriores sobre el tema. Sus investigaciones pioneras de la anatomía cerebral eran audaces e imaginativas, pero sus herramientas y métodos no podían explorar más que una minúscula parte de las complejidades que presentó, y las únicas metáforas disponibles (cuerdas y poleas y fluidos pasando por tuberías) eran demasiado toscas para suministrar a su imaginación una medida de las posibilidades de un modelo materialista del cerebro como mente. Así pues, no podemos reprocharle que llegara a la conclusión de que la mente, de la cual tanto conocía «desde el interior», tuviera que ser alguna otra cosa, una cosa pensante (res cogitans) no material. Empezó con mal pie asumiendo el «punto de vista en primera persona» como su acceso epistémico directo e incluso infalible a la conciencia, un paso que le encadenó a una ilusión de usuario que distorsionó sistemáticamente su investigación desde buen principio. Pero ¿qué más podía hacer? Observar el tejido cerebral aportaba una información absurdamente ridícula en comparación con una reflexión sobre sus pensamientos, sobre las sensaciones y las percepciones de las que disfrutaba o que aborrecía, sobre los planes que concebía y sobre las emociones que alteraban su estado de ánimo.

Desde entonces, filósofos, psicólogos y otros científicos se han basado en gran medida en la introspección como, al menos, una fuente abundante de pistas (y problemas), aplazando la pregunta de cómo es posible este maravilloso tesoro. Al fin y al cabo, es «evidente por sí mismo»; nuestras mentes conscientes están repletas de «ideas», «sensaciones» y «emociones», de las cuales tenemos un «conocimiento por familiaridad» que supera en intimidad e incorregibilidad a cualquier otro tipo de conocimiento (algo en lo que están de acuerdo muchos pensadores). La primacía de la «experiencia en primera persona» ha estado implícita en las prácticas de la mayoría de investigadores, si no un axioma declarado. A veces incluso se ha proclamado como sabiduría metodológica fundamental; John Searle (1980) lo establece categóricamente: «Recordad, en estas discusiones insistid siempre en el punto de vista en primera persona. El primer paso en el juego de manos operacionalista se da cuando intentamos determinar cómo sabríamos cómo serían las cosas para los otros» (p. 451).[3] En efecto, para muchos filósofos, el problema central no ha sido cómo ofrecer una descripción científica de la experiencia consciente, sino cómo retirar el «velo de la percepción» y pasar del «aquí dentro» al «mundo externo», una exploración iniciada con las Meditaciones de Descartes.

El precio que hay que pagar por seguir el consejo de Searle es llegar a todos tus fenómenos, los sucesos y las cosas que tiene que explicar tu teoría, a través de un canal que no está diseñado para la investigación científica, sino para el uso apresurado, sencillo y sin sofisticaciones de la vida cotidiana, brusca, polvorienta y aceitosa. Podemos aprender mucho acerca de cómo lo hace el cerebro (al fin y al cabo, también podemos aprender mucho acerca de los ordenadores insistiendo siempre en el punto de vista del escritorio), pero solo si tenemos presente que nuestro canal está sometido a una hipersimplificación sistemática y es siempre metafórico, no literal. Esto significa que debemos resistirnos a la seductora tentación de postular una panoplia de propiedades subjetivas especiales (llamadas habitualmente qualia) a las cuales nosotros, y solo nosotros, tenemos acceso. Son buenos elementos para nuestra imagen manifiesta, pero deben «agruparse», tal como dicen los fenomenólogos, cuando pasamos a la explicación científica. La incapacidad de apreciar esto conduce a una henchida lista de cosas que tienen que explicarse y que incluye, de forma destacada, una Pregunta Difícil que no es nada más que un artefacto de la incapacidad de reconocer que la evolución nos ha dado un don que sacrifica la verdad literal a cambio de utilidad.

Imaginemos que pedimos un consejo y nos dicen «¡Usa el páncreas!» o «¡Usa el hígado!». Sin duda, no tendríamos ni idea de qué acción emprender. Y cuando un maestro nos apremia a «usar el cerebro» nos quedaríamos bloqueados si no lo interpretáramos como «¡Usa tu mente!», esa cosa pensante con la que estamos tan íntimamente familiarizados que apenas es distinguible de nosotros, de nuestro yo. No es sorprendente que nos resistamos a considerarla una ilusión: si es una ilusión, ¡también lo somos nosotros!

Si nosotros, nuestro yo, no fueran más que parte de las ilusiones de usuario de los demás, ¿no llegaríamos a la conclusión de que, en realidad, la vida no tiene sentido alguno? No. La imagen manifiesta que ha sido conjuntada por los procesos evolutivos genéticos a lo largo de miles de millones de años, y por los procesos evolutivos culturales en miles de años, es un sistema extremadamente sofisticado de útiles representaciones metafóricas de la realidad subyacente revelada en la imagen científica. Se trata de una ilusión de usuario que usamos con tal pericia que la tomamos por la realidad descarnada, cuando, de hecho, tiene varias capas de barniz interpretativo. La imagen manifiesta forma nuestro Umwelt, el mundo en que vivimos para casi todos los propósitos humanos (ciencia aparte). Aprendemos cosas sobre la realidad a través de categorías de colores, sonidos, aromas, objetos sólidos, ocasos y arco iris, personas y sus intenciones, promesas, amenazas y garantías, instituciones y artefactos. Observamos nuestras perspectivas, tomamos nuestras decisiones, planificamos nuestras vidas y comprometemos nuestros futuros en sus términos, y esto es lo que hace que la imagen manifiesta sea relevante… para nosotros. Es un asunto de vida o muerte, ¿qué podría ser más relevante? Nuestras propias reflexiones sobre todo esto por fuerza están formuladas en términos de significados, o contenidos, el único «acceso» fácilmente utilizable que tenemos a lo que sucede dentro de nuestras cabezas.

Mientras Searle ha remarcado durante años la importancia de adoptar un punto de vista en primera persona (¿Qué veo? ¿Ser yo es ser como qué?), un filósofo que ha valorado muy pronto las «virtudes antisépticas» de adoptar un punto de vista en tercera persona (¿Qué quiere? ¿De qué son conscientes?) es Jonathan Bennett, cuya pequeña monografía Rationality (1964) se propuso estudiar indirectamente la racionalidad humana estudiando la no racionalidad de las abejas. Al insistir en adoptar el punto de vista en tercera persona, y empezar con una criatura humilde

pero de una competencia asombrosa, Bennett minimiza la tentación de caer en la inevitable práctica de la identificación por el contenido que es el sello de los métodos introspectivos.

Lo que quiero decir es lo siguiente: si quieres hablar sobre tus estados mentales, debes identificarlos por su contenido. «¿Qué idea? Mi idea de caballo. ¿Qué sensación? Mi sensación de blanco». ¿Cómo si no? No hay forma de que puedas identificar tus estados mentales «desde dentro» como, pongamos por caso, el «concepto J47» o la «sensación de color 294». Al dar por sentado el contenido de tus estados mentales, escogiéndolos por su contenido, estás barriendo bajo la alfombra todos los problemas de indeterminación o vaguedad del contenido. Leer tu propia mente es demasiado fácil; leer la mente de una abeja coloca los problemas bajo los focos. No tendremos una ciencia completa de la conciencia hasta que podamos adaptar nuestras identificaciones de estados mentales por su contenido, en la imagen manifiesta, con las identificaciones de estructuras y sucesos subpersonales de información, en la imagen científica, que son responsables causalmente de generar los detalles de la ilusión de usuario en la que operamos.

He aquí otra fuente de la persistencia del punto de vista cartesiano. Al suponer que nosotros, personas normales, somos racionales y, en consecuencia, tenemos comprensión (no solo competencia), aceptamos tácitamente nuestro uso diario del enfoque intencional no solo como algo práctico y valioso, sino como la realidad pura y dura de la mente humana. Esto nos sitúa en una distinguida compañía: somos diseñadores inteligentes, como el Diseñador Inteligente que nos diseñó. ¿Quién no querría tal honor? De este modo solemos conceder, a nosotros y a nuestros colegas, más crédito por la autoría de nuestras creaciones y más culpa por nuestras maldades que el que estaría justificado a partir de una perspectiva lisa y llana de la causalidad implicada.

Además (y aquí es cuando tenemos que pagar el precio), el punto de vista cartesiano encaja muy bien, según parece, con las ideas tradicionales de libre albedrío y responsabilidad moral. Hace unos años me di cuenta de la profundidad de esta impresión cuando intentaba dejar al descubierto los fundamentos ocultos de la resistencia a cualquier versión de una descripción de la conciencia como la que aquí he bosquejado, y descubrí que muchos científicos cognitivos (y no solo los profanos) eran renuentes a ni tan solo tener en cuenta esta doctrina. Tras

sofocar diversas de sus objeciones, terminaban confesando: «Pero ¿qué sucede con el libre albedrío? ¿No mostraría una descripción totalmente materialista de la conciencia que no podemos tener ninguna responsabilidad moral?». La respuesta es no, en absoluto, y voy a explicar el porqué en pocas palabras (ya he presentado mi postura en dos libros y muchos artículos, por lo que seré breve). El punto de vista tradicional del libre albedrío, como un poder personal por alguna razón aislado de la causación física, es incoherente e innecesario como fundamento para el sentido y la responsabilidad moral. Los científicos y filósofos que afirman que el libre albedrío es una ficción, una ilusión, están en lo cierto: forma parte de la ilusión de usuario de la imagen manifiesta. Así, el libre albedrío se sitúa en la misma categoría que los colores, las oportunidades, los dólares, las promesas y el amor (unos pocos ejemplos relevantes de un gran conjunto de ofrecimientos); si el libre albedrío es una ilusión, estos otros ofrecimientos también lo son, y por las mismas razones. No es una ilusión que querríamos eliminar ni desarmar; es donde vivimos, y sin ella no podríamos vivir de la manera en que lo hacemos. Pero cuando estos científicos y filósofos pasan a afirmar que su «descubrimiento» de esta ilusión (benigna) tiene implicaciones importantes para la ley, para saber si somos responsables o no de nuestras acciones y creaciones, sus argumentos se esfuman. En efecto, deberíamos quitarnos de encima los hirientes arreos del retribucionismo, que responsabiliza de modo absoluto a las personas por sus obras (a los ojos de Dios); en su lugar, deberíamos establecer un sistema de moralidad y justicia sensato, práctico y defendible, que castigue cuando sea necesario castigar, pero con un marco o postura muy diferente. Podemos captar la idea preguntándonos: si, dado que el libre albedrío es una ilusión, nadie es nunca responsable de lo que hace, ¿deberíamos eliminar las tarjetas rojas y amarillas del fútbol, las faltas técnicas del baloncesto y todos los demás sistemas de penalización de los deportes?

Los fenómenos del libre albedrío y la responsabilidad moral, valiosos elementos en la ontología de la imagen manifiesta humana, sobreviven con salud de hierro después de extraerles parte de la magia acumulada de la tradición y reformularlos en términos de la realidad científica. Con independencia de si estoy en lo correcto con mi afirmación de que los fenómenos del libre albedrío y la responsabilidad, adecuadamente replanteados y entendidos, son elementos defendibles de nuestra ontología más seria, debemos admitir que el miedo a que estas

trascendentales características de la vida cotidiana puedan quedar condenadas a desaparecer genera una poderosa resistencia que distorsiona la imaginación de las personas que intentan descubrir qué es la conciencia humana.

Nicholas Humphrey es el escritor que con más vehemencia ha llamado la atención sobre el prejuicio indiscutido que tienen las personas a favor de descripciones «espirituales» en su obra Soul Searching: Human Nature and Supernatural Belief (1995). Tal como Humphrey explica, las personas tienden a considerar la creencia en lo sobrenatural no solo aceptable, sino también moralmente loable. La credulidad va de la mano de la devoción. Muchos creen que la mente humana es el último bastión de todo lo que hay de sagrado en este mundo, y explicarla implicaría destruirlo; así, para que no salte todo por los aires, es mejor declarar que la conciencia está fuera del alcance de la ciencia. Como ya hemos visto, hay una enorme resistencia emocional contra cualquier consideración que pueda parecer que pone en duda la suposición de que los animales no humanos (el oso pardo, el perro, el delfín) tienen mentes que, si bien no son como las nuestras, sí que son lo bastante parecidas como para proporcionarles una cierta postura moral, quizá no responsabilidad moral como tal, pero al menos el derecho de no ser maltratados.

Por todas estas razones, resistirse a la fuerza de gravedad cartesiana exige un arduo esfuerzo de la imaginación, a la vez que debemos evitar pasarnos de largo en nuestra resistencia. Tenemos que dejar de lado algunas intuiciones que parecen casi indudables y tomar en serio algunas sugerencias que, a primera vista, se antojan paradójicas. No es una tarea fácil, pero la ciencia nos ha mostrado cómo hacerlo una y otra vez. Los niños en edad escolar pueden lanzar por la borda las intuiciones precopernicanas y pregalileanas sin inmutarse, y cuando ya son adolescentes, pueden sustituir algunas de sus intuiciones newtonianas por intuiciones einsteinianas, sin demasiados problemas. Sentirse cómodo con la física cuántica es todavía una tarea pendiente (al menos para mí, lo reconozco, a pesar de todos los ejercicios mentales que he hecho); en mi caso, es más fácil aceptar las extrañas inversiones del razonamiento de Darwin, de Turing y de Hume. Al presentar un esbozo de las causas de la gravedad cartesiana, espero haber ayudado a todos aquellos que no están muy convencidos a encontrar un punto de vista desde el cual diagnosticar sus propios fracasos de la imaginación y, así, poder superarlos.

La conciencia humana es diferente a todos los otros tipos de conciencia animal, en el sentido de que, en buena parte, es un producto de la evolución cultural, que instala una abundancia de palabras y demás recursos de pensamiento en nuestros cerebros, con lo que crea una arquitectura cognitiva muy diferente a las mentes «de abajo arriba» de los animales. Al proporcionar a nuestras mentes sistemas de representaciones, esta arquitectura nos da una perspectiva (una ilusión de usuario) desde la cual tenemos un acceso limitado y sesgado a los mecanismos de nuestros cerebros, que sin querer interpretamos erróneamente como una representación (proyectada sobre el mundo externo, sobre una pantalla o escenario privado) de las propiedades externas del mundo (colores, olores, sonidos…) y de muchas de nuestras respuestas internas (expectativas cumplidas, deseos identificados…). El inacabable torrente de mecanismos de autoobservación y de reflexión en que nos enfrascamos durante nuestra vida en vigilia es lo que nos permite, solo a nosotros, comprender nuestras competencias y muchas de las razones por las cuales el mundo es como es. Gracias a esta infestación de estructuras de información simbióticas producto de la evolución cultural, nuestros cerebros están facultados para ser diseñadores inteligentes, de artefactos y de nuestras propias vidas.

******** En inglés, figment, para hacer un juego de palabras con el pigment de la pintura. (N. del t.)

******** Se puede visualizar la imagen de la figura utilizando un lector de códigos QR o bien accediendo a http://pasadopresente.com/images/PDF-LIBROS/delasbacteriasabach_imagenesqr.pdf

15

LA ERA DEL POSDISEÑO INTELIGENTE

¿QUÉ LÍMITES TIENE NUESTRA COMPRENSIÓN?

Si el cerebro fuera tan simple para poderlo comprender, nosotros seríamos tan simples que no podríamos.

E

MERSON P. PUGH. The Biological Origins

of Human Values

Desde los tiempos prehistóricos, la comprensión humana ha ido creciendo de modo constante. Durante más de cuarenta milenios hemos vivido en la era del diseño inteligente: hemos elaborado vasijas, herramientas, armas, vestidos, viviendas y vehículos; hemos compuesto música y poemas; hemos inventado y perfeccionado prácticas agrícolas; hemos organizado ejércitos; todo con una mezcla de respeto a la tradición, improvisación imprudente y oportunista y una I+D deliberada, intencional y sistemática, interrumpida de vez en cuando por episodios de «inspirada» genialidad. Aplaudimos el diseño inteligente en todos los campos y desde la más tierna infancia aspiramos a lograr reconocimiento por nuestras creaciones. Entre todos los artefactos que hemos creado, tenemos el concepto de dios, el Diseñador Inteligente concebido a nuestra imagen y semejanza. ¿Qué mejor prueba del gran valor que otorgamos a los diseñadores inteligentes en nuestras sociedades?

Reconocemos el valor de estos frutos de nuestros esfuerzos, y nuestras leyes y tradiciones han sido diseñadas para crear un entorno artificial en que podamos conservar y mejorar nuestra riqueza acumulada. Es un entorno perfectamente real, no un mundo meramente virtual, pero no por ello es menos artificioso; es lo que denominamos civilización. Somos muy conscientes de que nuestra especie no tiene ninguna inmunidad especial a su propia extinción, como cualquier otra especie, y de que podemos desaparecer en una gran epidemia o catástrofe tecnológica o (una perspectiva solo un poco menos funesta) destruir la civilización y volver al «estado de naturaleza», como Hobbes lo denominaba, en que la vida es salvaje, brutal, sucia y breve. Pero ¿se nos ha pasado alguna vez por la cabeza que esta época de héroes con comprensión podría finalizar sin más, mientras Homo sapiens continúa reproduciéndose y sobrevive a su propio nombre, el hominino sapiente? Hay señales inquietantes de que nos estamos «hipercivilizando», gracias al ingenio de todos los inventos que nos ahorran trabajo y de los cuales nos hemos vuelto tan dependientes, y estamos entrando en la era del posdiseño inteligente.

La cita con que iniciaba este apartado, la ingeniosa reflexión de Pugh sobre el audaz proyecto de usar nuestros cerebros para entender nuestros cerebros se ha atribuido a muchos autores, en versiones diferentes, y es posible que se haya reinventado unas cuantas veces. Una variante es uno de mis comentarios favoritos de George Carlin:

Durante años he pensado que mi cerebro era el órgano más importante del cuerpo, hasta que un día pensé: ¡eh!, un momento, pero ¿quién me está diciendo esto?

¿Hay quizá aquí una verdad que nos está acechando o no es más que otra forma en que la gravedad cartesiana nos quiere desviar de nuestra búsqueda para entender la conciencia humana? Noam Chomsky (1975) ha propuesto una distinción que ha acaparado mucha atención y ha convertido a algunos discípulos: por un lado, tenemos problemas, que podemos resolver, y por el otro, misterios, que no podemos resolver. La ciencia y la tecnología han solucionado muchísimos problemas sobre la materia, la energía, la gravedad, la electricidad, la fotosíntesis y el ADN y han determinado las causas de las mareas, la tuberculosis, la inflación y

el cambio climático, por ejemplo; y continuamente se están haciendo progresos en miles de otros problemas. Pero por muy avanzada que sea nuestra capacidad de resolver problemas, algunos están fuera del alcance de la comprensión humana, y por ello mejor llamarlos misterios. La conciencia es el primero de la lista de Chomsky, junto con el libre albedrío. Algunos pensadores (que podemos denominar misterianos) han aceptado con entusiasmo esta afirmación injustificada basándose en la autoridad de Chomsky. Es posible que haya muchos misterios que queden sistemáticamente fuera del entendimiento humano, ahora y por siempre jamás, pero el argumento a favor de esta conclusión descorazonadora planteado por Chomsky y otros misterianos, aunque tiene un cierto atractivo, no es convincente. Tenemos aquí una representación del argumento del cierre cognitivo, a partir de diversas versiones:

Es un hecho innegable de la biología que nuestros cerebros son limitados, como los de todas las otras criaturas. A partir de nuestro punto de vista comparativamente olímpico, podemos ver que los peces son inteligentes a su modo, pero está claro que no están pertrechados para entender la tectónica de placas, mientras que los perros están en la inopia por lo que respecta al concepto de democracia. Cualquier cerebro está sometido al cierre cognitivo (McGinn 1990) con respecto a una serie de problemas que están más allá de sus posibilidades, inconcebibles e incomprensibles. No tenemos una res cogitans milagrosa en nuestra cabeza, solo una gran cantidad de tejido cerebral sometido a las leyes de la física y la biología.

Hasta aquí, todo en orden; no tengo ninguna objeción a este planteamiento, que expresa con claridad unos hechos innegables del mundo físico. Pero luego, prosigue el argumento:

Sería profundamente abiológico (¿no es cierto?) suponer que, de algún modo, nuestros cerebros humanos están exentos de estos límites naturales. Estas ilusiones de grandeza son reliquias obsoletas de nuestro pasado precientífico.

Esto resultaría convincente si no fuera por el hecho biológico, igualmente obvio, de que los cerebros humanos han sido equipados con miles de complementos, miles de recursos de pensamiento que multiplican su potencia cognitiva en varios órdenes de magnitud. El lenguaje, como ya hemos visto, es un invento clave, que amplía nuestros poderes cognitivos personales gracias a un medio que permite unirlos con todos los poderes cognitivos de todos los demás seres humanos inteligentes que piensan y que han pensado. El más listo de los chimpancés nunca puede comparar sus pensamientos con otros chimpancés de su grupo, y mucho menos con los millones de chimpancés que han vivido antes que él.

El punto débil del argumento del cierre cognitivo es la permanente dificultad de hallar buenos ejemplos de misterios. En cuanto planteas una cuestión que según tú nunca se podrá responder, pones en marcha el proceso que bien puede demostrar que estás equivocado: planteas un tema de investigación. Aunque tu cuestión pueda empezar con mal pie, como algo imposible de solucionar, es muy probable que esto quede desvelado por el proceso de intentar responderla. La curiosidad reflexiva de la filosofía (que metaanaliza cualquier pregunta planteada) es una garantía casi absoluta de que se producirá algo muy parecido a una búsqueda exhaustiva (a veces casi al azar, a veces dirigida con brillantez) de variaciones sobre la cuestión que pueden resultar más inteligibles. Plantear preguntas cada vez mejores es un punto clave para refinar y perfilar nuestra búsqueda de soluciones a los «misterios», y este refinamiento está mucho más allá de las capacidades de cualquier criatura carente de lenguaje. «¿Qué es la democracia?» Un perro nunca encontrará la respuesta, sin duda, pero es que nunca llegará a entender la pregunta. Nosotros podemos comprender las preguntas, lo que cambia radicalmente nuestra búsqueda, convirtiendo misterios insondables en problemas que vale la pena intentar resolver.

Quizá teniendo en cuenta este poder casi ilimitado del lenguaje para ampliar nuestro entendimiento, Chomsky ha moderado algo su postura (2014). Aunque, dice, hay una «distinción conceptual» entre problemas y misterios, «aceptamos las mejores explicaciones que nos ofrece la ciencia», incluso aunque no podamos imaginar cómo funcionan. «Ya no importa lo que podemos concebir. Hemos desistido de esto.» Dicho de

otro modo, gracias al lenguaje y a los recursos que la ciencia nos ofrece, podemos elaborar una buena teoría científica de algún fenómeno desconcertante, una teoría que vale la pena avalar, aunque en realidad no la comprendamos. Es decir, podemos justificar su aceptación e incluso podemos apostar nuestra propia vida en la pertinencia de sus predicciones, aunque no entendamos cómo ni por qué funciona. Tanto si esta revisión es del gusto de los misterianos como si no, es una idea interesante. Pero ¿puede ser cierta?

Descargar miles de recursos de pensamiento adquiridos culturalmente nos puede permitir amplificar radicalmente nuestros poderes, pero ¿no estaremos aplazando el punto en que se produce el cierre cognitivo? ¿Cuántos conocimientos puede absorber una mente o cerebro individual? Aquí aparece una ambigüedad en la suposición misteriana: ¿qué se está afirmando? ¿Que hay misterios que ninguna mente humana individual puede comprender? ¿O que hay misterios inalcanzables para la comprensión conjunta de civilizaciones enteras? La idea de la comprensión distribuida o compartida (la idea de que, como grupo, podemos comprender cosas que ningún individuo puede por sí solo) resulta absurda para muchas personas, fieles al ideal del bricolaje del diseñador inteligente, el genio que todo lo soluciona. Es un tema con muchas variaciones familiares. Una pintura de la escuela de Rembrandt es menos valiosa que una pintura del propio Rembrandt. Las novelas suelen estar escritas en solitario (la dura tarea de los editores que revisan y reorganizan el penúltimo borrador no suele reconocerse), y cuando equipos creativos tienen éxito (Gilbert y Sullivan, Rodgers y Hammerstein) casi siempre se trata de una división del trabajo: uno escribe la letra y el otro compone la música, por ejemplo. Pero fuera del campo de la ficción, la autoría múltiple ha sido habitual durante siglos, y en el mundo científico actual, hay muchos campos en que artículos con un único autor son una auténtica rareza.

Uno de los trabajos fundadores de la ciencia cognitiva, Planes y estructura de la conducta (1960), fue escrito por George Miller, Eugene Galanter y Karl Pribram. Su introducción de la idea de una unidad TOTE (Test-Operate-Test-Exit, ‘probar-operar-probar-salir’) fue una temprana formalización de la idea de bucles de retroalimentación, y representó un papel destacado en la transición del conductismo a la modelización cognitiva. A pesar de su gran influencia inicial, es una obra apenas leída en la actualidad, y una broma que solía oírse era que Miller la escribió,

Galanter la formalizó y Pribram se la creyó. La idea de que una división del trabajo de este tipo podía ser posible, y satisfactoria, resultaba chistosa en esa época, pero ya no. La ciencia está repleta de colaboraciones en que los teóricos (que comprenden el aparato matemático) y los experimentalistas e investigadores sobre el terreno (que se basan en los teóricos sin dominar la estructura matemática) trabajan juntos para crear trabajos con autoría múltiple en que muchos detalles solo son entendidos parcialmente por cada uno de los autores. Combinaciones así de comprensión especializada florecen en muchas otras disciplinas.

Imaginemos, pues, un libro con muchos autores y muchos volúmenes, La teoría científica de la conciencia, que es aceptado sin discusión por la comunidad científica. Digamos que los diversos volúmenes se convierten en los libros de texto estándar sobre la conciencia humana, empleados en todos los cursos de neurociencia, psicología, filosofía y otros campos en que la conciencia es un fenómeno relevante, y aunque algunos valientes afirman haber leído toda la obra, nadie se atreve a decir que domina todos los niveles explicativos. ¿Vindicaría esto el misterianismo chomskiano (la conciencia sigue siendo un misterio, pues ningún individuo puede concebirla en su totalidad)? ¿O derrocaría otro más de los candidatos misterianos a «misterio insondable»?

A medida que la civilización ha progresado, hemos aprendido que la división del trabajo permite hacer cosas. Una sola persona, o una sola familia, puede construir una casa sencilla o una canoa y una pequeña comunidad puede erigir un granero o una empalizada, pero se necesitan centenares de trabajadores con decenas de talentos diferentes para construir una catedral o un gran velero. Actualmente los artículos con revisión por expertos y centenares de coautores son la tónica general en instituciones como el CERN y otras instituciones de la Gran Ciencia (la Big Science). En muchos casos, ninguno de los autores puede afirmar tener más que una somera comprensión de toda la empresa; hemos llegado a un punto en que incluso los más brillantes y solitarios pensadores dependen de sus colegas para comentar, revisar y confirmar.

Pongamos el ejemplo de Andrew Wiles, el brillante matemático de Princeton que en 1995 demostró el Último Teorema de Fermat, un logro excelso de la historia de las matemáticas. Una mirada atenta al proceso por el que llegó al resultado, incluyendo las salidas en falso y las lagunas

inadvertidas en su primera versión de la demostración, demuestra que este logro fue, en realidad, la obra de muchas mentes, una comunidad de expertos en contacto, colaborando y compitiendo entre sí para obtener la gloria; y está claro que sin los incontables ladrillos matemáticos bien comprobados de los que dependía la demostración de Wiles, hubiera sido imposible, para Wiles o para cualquier otro, determinar que, en efecto, el teorema había quedado demostrado.[1] Si eres un lobo solitario y crees que has demostrado el Último Teorema de Fermat, te enfrentarás a este dilema: «o he demostrado el Último Teorema de Fermat o me estoy volviendo loco», y como la historia muestra que muchos matemáticos brillantes se han engañado pensando que habían tenido éxito, tendrás que tomarte muy en serio la segunda posibilidad; solo la aceptación formal y la consiguiente felicitación de tus colegas podrían calmar esta ansiedad.

Incluso los artistas, poetas y músicos, apreciados por su individualidad e «inspiración divina», dan lo mejor de sí cuando están bien familiarizados con las obras de sus predecesores. Los rebeldes del siglo XX, obsesionados por desafiar «los cánones establecidos» y por crear obras hiperoriginales, o bien han caído en el olvido o bien han demostrado que la fama de sus creaciones se debe a una mayor apreciación de las tradiciones de lo que les gustaría admitir. En cierta ocasión, el pintor Philip Guston admitió con elocuencia su dependencia indirecta de todo lo que había extraído y procesado de los diseños inteligentes de otros:

Creo que fue John Cage quien una vez me dijo «cuando empiezas a trabajar, todo el mundo está en tu estudio: el pasado, tus amigos, tus enemigos, el mundo del arte y, por encima de todo, tus propias ideas; todos están ahí. Pero cuando sigues pintando, empiezan a marcharse, uno tras otro, hasta quedarte totalmente solo. En ese momento, si tienes suerte, incluso tú mismo te marchas». (2011, p. 30)

¿Qué límites tienen los cerebros con los que nacemos? De momento, podemos fijarnos en que tanto si los límites son prácticos como absolutos, hemos hallado y perfeccionado un rodeo que aplaza el enfrentamiento con nuestra fragilidad: la colaboración, sistemática o informal. Los grupos pueden hacer cosas y (suponemos) comprender cosas que los individuos solos no pueden, y buena parte de nuestro poder

se deriva de este descubrimiento. Es posible resistirse a esta idea de comprensión grupal, pero solo (tal como yo lo veo) si elevamos la comprensión hasta unas alturas místicas que poco tiene que ver con la comprensión en la que nos basamos (en nosotros y en los demás) para solucionar nuestros problemas y crear nuestras obras maestras. Esto suaviza los bordes del argumento misteriano. Al ignorar el poder de la comprensión colaborativa, plantea un problema obsoleto considerando la comprensión como un don «a todo o nada», que se tiene al completo o no se tiene en absoluto, cuando bien pocas veces es así.

En su época, Descartes estaba muy preocupado por garantizar una comprensión perfecta para sus ideas «claras y nítidas» y, para ello, argumentaba, necesitaba demostrar la existencia de un Dios benigno, omnipotente y no engañoso. Su experimento mental planteaba que, por lo demás, habría un demonio maligno dedicado a engañarle acerca de sus convicciones más firmes, y esta «posibilidad en principio» fijó su método con seguridad (y también le ató de manos). Para Descartes, solo el tipo de certeza que reservamos para las verdades matemáticas evidentes (2 + 2 = 4, un triángulo en el plano tiene tres lados rectos y sus ángulos interiores suman 180°) era lo bastante buena para considerarse conocimiento real, y solo la comprensión cristalina que podemos tener de cada uno de los pasos individuales de una demostración simple en extremo pueden considerarse comprensión perfecta. Mientras que Descartes confiaba en Dios como garante de sus demostraciones, en la actualidad nos basamos en la improbabilidad de que varios pensadores lleguen, por rutas diferentes, al mismo resultado erróneo (fijémonos que se trata de una aplicación del mismo principio que dictaba llevar tres cronómetros en el barco, de modo que cuando la hora indicada empezaba a diferir, el que se desviaba era seguramente el erróneo). Tendemos a pasar por alto la importancia de disponer de una voluminosa experiencia de muchas personas que, de modo independiente, llegan a la misma respuesta para preguntas, por ejemplo, de multiplicaciones y divisiones, pero si esta no fuera nuestra experiencia, ninguna reflexión analítica sobre la necesidad intrínseca de las matemáticas (o de la existencia de un Dios misericordioso), por muy profunda que fuera, nos convencería de confiar en nuestros cálculos. ¿Es la aritmética un sistema satisfactorio de cálculo? Probablemente, con una probabilidad tan alta que puedes confiar tu vida en él sin preocuparte en lo más mínimo.

«MIRA, MAMÁ… ¡SIN MANOS!»

La civilización avanza ampliando el número de operaciones importantes que podemos efectuar sin pensar en ellas.

A

LFRED NORTH WHITEHEAD

No entiendo lo que no puedo crear.

R

ICHARD FEYNMAN

He sostenido que la I+D básica, ciega y de abajo arriba que efectúa la selección natural ha ido creando grúas (productos que permiten ahorrar trabajo y hacen más eficaz el trabajo de diseño) que han abierto nuevas zonas del espacio de diseño para grúas adicionales, en un zoom acelerado hacia la era del diseño inteligente, en la que puede prosperar la I+D dirigida, sistemática, basada en razones, reflexiva y de arriba abajo. Este proceso ha tenido éxito al cambiar el equilibrio de fuerzas selectivas que nos conforman, a nosotros y todos los demás organismos, y al crear teorías de alta capacidad predictiva que explican retrospectivamente los procesos de su propia creación. Esta cascada de grúas no es un milagro, no es un don divino, sino un producto natural del proceso evolutivo fundamental, junto con los otros frutos del árbol de la vida.

Repasemos este proceso. A lo largo de miles de años, los seres humanos hemos llegado a apreciar las capacidades de las mentes individuales. A partir de los hábitos instintivos de todos los seres vivos, sabemos distinguir el alimento de las sustancias venenosas y, al igual que otros organismos locomotores, somos muy sensibles a la animacidad (el movimiento guiado) de otros seres móviles y, especialmente, a las creencias y deseos (información y objetivos) que guían tales

movimientos, e intentamos controlar, lo mejor que podemos, quién sabe qué y quién quiere qué, para dirigir nuestros propios esfuerzos en este juego del escondite. Este sesgo nativo es la base genética del enfoque intencional, nuestra práctica de tratar a los demás como agentes racionales guiados, en su mayor parte, por creencias ciertas y por deseos bien planteados. Nuestro interés continuo en estos temas ha generado la psicología popular en la que confiamos para ser coherentes unos con otros. La usamos para predecir y explicar no solo los comportamientos repetitivos que observamos en nuestros vecinos y en nosotros mismos, así como los «movimientos forzados» que resultaría estúpido no ejecutar, sino también los arrebatos de genialidad que son el sello de los genios. De este modo, nuestras expectativas se ven confirmadas muy a menudo, lo que refuerza nuestra fidelidad al enfoque intencional. Cuando nuestras expectativas se ven defraudadas, tendemos a apoyarnos en «explicaciones» de nuestros fallos que, en el mejor de los casos, son acertadas conjeturas, y en el peor, engañosa mitología.

Alentamos a nuestros hijos a ser curiosos y creativos e identificamos conscientemente los surcos y las fronteras de nuestros procesos de pensamiento para intentar superarlas. Las mentes que más valoramos son las que no son ni demasiado predecibles (aburridas, poco estimulantes) ni demasiado caóticas. La práctica lleva a la perfección, y hemos inventado juegos que nos animan a practicar nuestras jugadas mentales, como el ajedrez, el go o el póquer, así como dispositivos protésicos (telescopios, mapas, calculadoras, relojes y miles de otros artefactos) que nos permiten aplicar nuestras jugadas mentales en entornos cada vez más artificiales y sofisticados. En cada ámbito de la investigación y el diseño, tenemos grupos de expertos muy organizados que colaboran para crear y perfeccionar teorías y demás productos, y hemos adoptado tradiciones y mecanismos comerciales para disponer del tiempo, la energía y los materiales necesarios para tales proyectos. Somos diseñadores inteligentes que habitamos en un mundo diseñado inteligentemente por nuestros ancestros para diseñadores inteligentes. Y ahora, tras siglos de soñar con esta perspectiva, hemos empezado a diseñar y fabricar artefactos que pueden diseñar y fabricar artefactos (que pueden diseñar y fabricar artefactos…).

«El trabajo compartido es más llevadero», un refrán que se puede aplicar tanto al trabajo mental como a construir un granero, pero ahora estamos descubriendo que el trabajo de diseño sin intervención directa suele ser

más sencillo y ligero y, gracias a los artefactos que hemos diseñado recientemente, más competente.

La nanotecnología, el boyante nuevo campo de la química y la ciencia de materiales que está empezando a construir artefactos átomo a átomo, parte del trabajo de brillantes y pacientes pioneros que han elaborado sofisticadas herramientas para manipular (mover, cortar, aislar, inmovilizar) fragmentos de materia a escala nanométrica (un nanómetro es la milmillonésima parte de un metro). Como la GOFAI de antaño, la nanotecnología empezó como un diseño inteligente de arriba abajo, un brillante método para hacer a mano grandes repertorios de «fármacos milagrosos», «materiales inteligentes» y otros nanorrobots. Ha cosechado muchos triunfos, y sin duda cosechará muchos más, especialmente con el nuevo nanorrecurso de la CRISPR (para una introducción breve y no técnica, véase Specter 2015). Como la PCR (reacción en cadena de la polimerasa), la técnica que revolucionó la secuenciación de los genes, la CRISPR, que permite editar, cortar y empalmar genes a voluntad, sustituye a técnicas muy sofisticadas y laboriosas; se trata de una invención que ahorra trabajo y, así, reduce en órdenes de magnitud el tiempo y los esfuerzos necesarios. Jennifer Doudna, de la Universidad de California en Berkeley y Emmanuelle Charpentier, ahora en el Instituto Max Planck, son dos de los diseñadores sumamente inteligentes de esta nueva grúa.

Todas estas técnicas, como las desarrolladas por Pixar y otras empresas de animación por ordenador, crean procesos automatizados que sustituyen a miles de días de brillante trabajo monótono (no es ningún contrasentido: se trata de personas con mucho talento haciendo un trabajo muy repetitivo pero muy exigente). Cuando Walt Disney Productions estrenó Blancanieves y los siete enanitos en 1937, sorprendió al mundo con sus realistas animaciones, el resultado del trabajo de cientos de hábiles animadores que trabajaban en equipos bien organizados para solucionar los problemas de obtener una acción vívida, con todas las irregularidades de la realidad, y plasmarla en miles de cels, o fotogramas. Todos estos heroicos talleres artísticos son ahora una reliquia histórica; el talento que se necesitaba para ser uno de esos animadores fotograma a fotograma está ya obsoleto, y lo mismo es cierto de los talentos de los primeros biólogos moleculares, que aislaban con agudeza fragmentos de genes y pacientemente les persuadían para que divulgaran sus secuencias, un codón tras otro. Podríamos explicar

historias similares de automatización de tareas que antes implicaban un tedioso trabajo intelectual, desde la astronomía al análisis textual. En general, estas tareas consisten en recopilar, ordenar y refinar datos a gran escala, lo que proporciona al interpretador humano de los datos más tiempo para reflexionar acerca de los resultados (nunca olvidaré el tiempo que pasé un día en el laboratorio de un prometedor joven neurocientífico, recopilando datos de experimentos con macacos que tenían electrodos permanentes implantados en el cerebro: más tarde, ese mismo día, le pregunté su punto de vista sobre una controversia teórica candente en esa época, acerca del papel que tenía la actividad en diversas áreas cerebrales para modular la conciencia; suspiró y contestó: «¡No tengo tiempo para pensar! Estoy demasiado ocupado con los experimentos»). Las nuevas técnicas que minimizan el brillante trabajo tedioso son extraordinariamente competentes, pero siguen siendo recursos (no colegas robóticos) que dependen del todo de las decisiones e intenciones de los directores y usuarios inteligentes de esos recursos, como los directores de laboratorios o los directorios de estudios de animación cinematográfica.

Sin embargo, actualmente estamos empezando a apreciar y a explotar la Segunda Regla de Orgel: la evolución es más lista que tú. En muchos campos, los diseñadores inteligentes están aprovechando los incansables algoritmos de abajo arriba de la selección natural (y sus parientes cercanos) para hacer el trabajo sucio de búsquedas masivas, el trabajo de hallar valiosísimas agujas en todo tipo de pajares.

En algunos casos, esta exploración implica la selección natural biológica real en el laboratorio. Por ejemplo, Frances Arnold, en el Caltech, emplea una laureada técnica de ingeniería de proteínas para crear nuevas proteínas, criándolas a partir de ADN con mutaciones al azar, en realidad; ha concebido sistemas para generar enormes poblaciones de variantes génicas (recetas de ADN para crear proteínas) y luego probar las proteínas obtenidas y así poder descubrir en ellas capacidades nunca antes presentes en la naturaleza.

Estamos elaborando nuevas herramientas para la ingeniería de proteínas y empleándolas para crear catalizadores nuevos y mejorados para la fijación de carbono, para la liberación de azúcar de polímeros renovables

como la celulosa y para la biosíntesis de combustibles y otros productos químicos. (Arnold 2013)

Arnold se dio cuenta de que, como el espacio de posibles proteínas es Vastamente más grande que el espacio de las proteínas existentes, sin duda habrá rutas practicables de evolución gradual que nunca se han explorado y que llevan a destinos que nos pueden proporcionar fármacos, tejidos y catalizadores extraordinarios, un ejército de nanorrobots que pueden servirnos una vez los encontremos. Cuando estaba haciendo el doctorado, un investigador superior le advirtió que no se conocía ninguna proteína con las propiedades que ella intentaba obtener; su atrevida respuesta fue «eso es porque nunca ha habido una selección para ellas».

En consecuencia, estas enzimas pueden abrirnos grandes regiones del «espacio químico» que no se podían explorar con los esfuerzos anteriores de la química médica. (Arnold 2013)

Frances Arnold ha creado una tecnología para crear nuevas proteínas (largas cadenas de aminoácidos que, cuando se unen, se pliegan formando avanzados nanorrobots evolucionados con poderes remarcables). David Cope, profesor emérito de música en la Universidad de California en Santa Cruz, ha desarrollado una tecnología notablemente diferente. Cope usa un programa informático para generar nueva música, largas secuencias de notas y acordes que, unidas, forman composiciones musicales con poderes notables: imitaciones de Bach, de Brahms, de Wagner, de Scott Joplin e incluso de canciones de comedias musicales (Coper y Hofstadter 2001). ¿Cuán originales son los miles de composiciones creadas por el EMI (Experiments in Musical Intelligence) de Cope? Está claro que son obras derivadas de otras obras previas e incluyen muchos préstamos de los grandes compositores cuyos estilos imita, pero no son meras copias, ni tampoco meras copias con unas pocas mutaciones al azar; son obras mucho mejores. El programa obtiene y digiere las obras de los maestros y, mediante sus procesos computacionales, extrae el fundamento, los aspectos esenciales, el estilo

de cada compositor para luego crear nuevas piezas con ese mismo estilo, lo que constituye un logro musical muy sofisticado. Si sois músicos, podéis intentar componer una obra para piano al estilo de Chopin o de Mozart, o de Count Basie o Erroll Garner; una caricatura o parodia sencilla no es muy difícil (especialmente de un pianista de jazz tan amanerado como Errol Garner), pero componer buena música exige una percepción y talento musical considerables… en un compositor humano.

Experiments in Musical Intelligence, diseñado y mejorado por Cope durante más de tres décadas, ha generado muchas piezas para piano, canciones, sinfonías y otras composiciones bien construidas, sin ninguna intervención de Cope, excepto por la valoración estética final acerca de cuáles de entre toda la multitud merecen escucharse. En el Festival Bach de Montreal de 2015, en el que di una charla en que se resumían los puntos principales de este libro, organicé una interesante prueba del EMI (una más de las muchas que se han hecho a lo largo de los años), que se cerraba con la interpretación, a cargo del pianista ucraniano Serhiy Salov, de cuatro piezas breves para piano. Dije a la audiencia, formada por más de 300 aficionados a Bach, que como mínimo una de las piezas era de Bach y una era del EMI; tras la interpretación, el público votó (con los ojos cerrados). Docenas de asistentes consideraron dos creaciones de EMI como auténticas obras de Bach (quizá no eran mayoría, pero se acercaba), y cuando pedí a los que habían acertado todas las piezas que se levantaran, solo lo hicieron unas doce personas, que fueron saludadas con un gran aplauso.

Cope, al igual que Arnold, establece los objetivos y decide cuándo cantar victoria, pero por lo demás, no interviene para nada. Estos dos proyectos de investigación, tan diferentes, son variaciones sobre el tema de Darwin de la selección metódica, en que la fuerza selectiva de la selección natural se canaliza a través del sistema nervioso de un agente previsor, exigente y con un propósito. Pero el trabajo duro de alzamiento se deja en manos de los implacables poderes de búsqueda de patrones de los algoritmos de la selección natural, en cascadas de ciclos de generación y prueba sin pizca de comprensión que poco a poco van perfeccionando el propio proceso de búsqueda. Como la selección natural es una familia de algoritmos que no depende del sustrato, se puede producir en cualquier medio que tenga unas pocas propiedades simples; la evolución in silicio (simulada en un programa informático) es muchas veces más rápida y barata que la evolución in vivo y se puede aplicar a casi cualquier

cuestión o problema. El reciente libro de Pedro Domingos, The Master Algorithm (2015), es un vívido y acreditado estudio de las nuevas variedades de sistemas darwinianos (y quizá podríamos decir también «darwiniescos») de «aprendizaje automático» o «aprendizaje profundo». Domingos simplifica toda esta avalancha clasificándola en cinco «tribus de aprendizaje automático»: los simbolistas (descendientes de la GOFAI), los conexionistas (descendientes de las neuronas lógicas de McCulloch y Pitts, véase capítulo 6), los evolucionarios (los algoritmos genéticos de John Holland y su prole), los bayesianos (que han concebido algoritmos prácticos para lograr las competencias de las redes jerárquicas de generadores bayesianos de expectativas) y los analogizadores (descendientes del algoritmo del vecino más cercano de Fix y Hodges [1951]). Las cinco familias reproducen, de diferentes maneras, el patrón de la selección natural. Al estar basados en sistemas informáticos, todos están formados, en último término, por las competencias sin comprensión más simples de Turing (aritmética y ramificación condicional) y, excepto quizá por las creaciones de los simbolistas, son repetitivas trituradoras buscadoras de agujas en pajares, sistemas de abajo arriba, que gradualmente y con gran fiabilidad se van acercando poco a poco a buenas respuestas (o por lo menos, lo bastante buenas) a problemas intimidantes.

John Holland, el apreciado y recientemente fallecido mentor de docenas de brillantes científicos cognitivos e informáticos en el Santa Fe Institute y en la Universidad de Michigan, inventó los algoritmos genéticos, en que los paralelismos con la evolución por selección natural resultan evidentes (y exquisitos para los darwinianos): se generan enormes poblaciones de variantes de codificaciones y a cada una de ellas se le da la oportunidad de avanzar en la resolución de un problema; los ganadores de esta prueba ambiental se reproducen (una reproducción completa, con «sexo» y «cruces» como en la mezcla aleatoria de genes que se produce en la creación de los espermatozoides y los óvulos). A lo largo de muchas generaciones, la competencia de las cadenas de código informático creadas inicialmente al azar se multiplica y se perfecciona. Los algoritmos genéticos se han usado para diseñar las fascinantes criaturas virtuales, producto de evolución, de Karl Sims (véanse los diversos sitios web dedicados a este serio terreno de juego de la imaginación) y también sensatos triunfos de la ingeniería como circuitos impresos y programas informáticos. Domingos señala (p. 133) que en 2005 se presentó una patente para un sistema de optimización de una

fábrica diseñado genéticamente (cuidado, general Groves, ¡le están pisando los talones!). Los arquitectos han empezado a usar algoritmos genéticos para optimizar las propiedades funcionales de los edificios (como su resistencia, seguridad, uso de materiales y aprovechamiento de la luz y la energía).

En la investigación científica, el aprendizaje automático se está aprovechando para solucionar, mediante fuerza bruta, problemas que están más allá de las capacidades analíticas humanas. Cabe señalar que Richard Feynman, por más que fuera un brillante físico teórico, pasó muchos de sus últimos días explorando el uso de superordenadores para resolver problemas de física que desafiaban su genialidad con las ecuaciones. Y vivió lo suficiente para ver más o menos refutada su máxima, con la que iniciábamos este apartado. Si bien aún puede ser cierto que no se puede comprender lo que no se puede crear, crear algo ya no es garantía de poder comprenderlo, como antes. Ahora se pueden crear cosas (de un modo muy indirecto) que hacen lo que queremos que hagan, pero sin poder entenderlas. A veces a esta situación se

denomina ciencia de la caja negra. Compras una caja negra de ultimísima tecnología, le proporcionas tus datos en bruto y ella te da el análisis; ahí tienes las gráficas, a punto para imprimir y publicar, pero no puedes explicar en detalle cómo funciona, ni repararla si se estropea, y tampoco está claro en absoluto que alguien pueda. Ahora bien, esta posibilidad no ha dejado de estar delante de nuestras narices; podemos controlar las cosas que hacemos «a mano» (barcos, puentes, motores, sinfonías) mientras las construimos y entendemos cada uno de los pasos del proceso, pero las cosas que hacemos «a la vieja usanza» (hijos, nietos…) desafían nuestra capacidad de comprensión a causa de nuestra inconsciencia de los procesos que las crean. Actualmente generamos creaciones hijas, creaciones nietas, creaciones biznietas… que dependen de procesos que no podemos seguir con detalle, incluso pudiendo demostrar que los resultados son válidos.

El uso de computadoras en la investigación ha llevado a diversos tipos de retos para la máxima de Feynman. Algunas demostraciones matemáticas realizadas (en parte o por completo) por ordenador son demasiado largas para que un matemático humano compruebe cada uno de los pasos, lo que, con razón, ha sido el criterio de aceptación durante miles de años. Un ejemplo famoso es la demostración con ayuda de un ordenador del

teorema de los cuatro colores en 1976, planteado por primera vez por Möbius en la década de 1840: cualquier mapa con regiones continuas, separadas por fronteras compartidas, se puede colorear con cuatro colores, de modo que no haya regiones adyacentes con un mismo color. Tras varias demostraciones fallidas por parte de algunos de los mejores matemáticos del mundo, Kenneth Appel y Wolfgang Haken aprovecharon la potencia de un ordenador para considerar y desestimar las casi 2.000 posibilidades que tenían que descartarse, tal como ellos habían demostrado. Durante varios años, esta demostración no fue aceptada porque parecía implicar una serie de pasos efectuados por la computadora imposibles de comprobar por un ser humano, pero actualmente hay un amplio consenso en considerar que se trata de un teorema demostrado (y desde entonces se han construido otras demostraciones, todas con ordenador). El teorema de los cuatro colores es un resultado «intuitivo»: nadie había conseguido producir un contraejemplo claro a pesar de personas que, en su conjunto, habían invertido siglos intentándolo, de modo que la mayoría de matemáticos asumieron que era cierto mucho antes de que fuera demostrado. Pero también hay teoremas antiintuitivos que han sido demostrados con ayuda de ordenadores. En el ajedrez, por ejemplo, los expertos siempre habían considerado que la regla de los cincuenta movimientos, que establece que una partida puede finalizar en tablas si en los últimos cincuenta movimientos no ha habido ningún movimiento de peón ni captura de pieza, era generosa con creces, pero fue tumbada por el descubrimiento, con análisis por ordenador, de algunas redes de mate (secuencias ganadoras de jugadas de las cuales no se puede salir una vez se han iniciado) sin capturas ni movimientos de peón que superaban los cincuenta movimientos. Tras algunos experimentos para modificar el número, la Federación Internacional de Ajedrez decidió oficialmente mantener la regla, ya que era una posibilidad que, en principio, nunca se daría en partidas serias entre humanos.

Los programas informáticos que analizan posiciones de ajedrez, igual que los que demuestran proposiciones matemáticas, son programas diseñados inteligentemente, de arriba abajo. Los programas de los que se ocupa Domingos son muy diferentes. Tal como lo expresa él mismo: «podemos ver el aprendizaje automático como lo opuesto a la programación, del mismo modo que una raíz cuadrada es la operación inversa a la elevación al cuadrado, o la integración es la operación inversa a la diferenciación» (p. 7). He aquí una más de las extrañas

inversiones del razonamiento o, mejor dicho, un ejemplo más de la inversión darwiniana básica: la competencia sin comprensión. La «hipótesis central» del libro de Domingos es de una audacia sin par:

Todo el conocimiento (pasado, presente y futuro) se puede obtener a partir de un único algoritmo de aprendizaje universal.

Es lo que denomino el Algoritmo Maestro. Si un algoritmo así es posible, inventarlo sería uno de los mayores logros científicos de la historia. De hecho, el Algoritmo Maestro sería la última cosa que tendríamos que inventar, ya que, una vez puesto en marcha, empezaría a inventar todo lo que se puede inventar. Todo lo que necesitamos es suministrarle la cantidad adecuada de datos relevantes y descubrirá el conocimiento correspondiente. (p. 25)

No está claro si es totalmente sincero, porque poco después se echa atrás:

Algunos admiten que el aprendizaje automático puede hallar regularidades estadísticas en los datos, pero dicen que nunca descubrirá nada profundo, como las leyes de Newton. Podemos decir que aún no, pero apuesto a que lo hará. (p. 39)

Una apuesta, pues, no una hipótesis que él crea que actualmente se puede fundamentar mediante argumentos razonados en su libro. En cualquier caso, resulta útil disponer de su articulación de esta hipótesis radical, ya que sin duda muchas personas casi tienen pesadillas sobre tales eventualidades y ayudará a arrojar algo de luz escéptica sobre ellas. Podemos empezar con la afirmación a la que Domingos responde con su apuesta: ¿podrá el aprendizaje automático avanzar más allá del proceso de hallar «regularidades estadísticas»? Domingos apuesta que sí, pero ¿qué base tiene su optimismo?

LA ESTRUCTURA DE UN AGENTE INTELIGENTE

Hemos visto que las redes bayesianas son excelentes para sonsacar las regularidades estadísticas relevantes para el organismo: sus ofrecimientos. Los cerebros animales, pertrechados por la selección natural con estas redes, pueden guiar a los seres en que habitan con una excelsa habilidad, pero por sí mismos tienen poca capacidad de adoptar nuevos puntos de vista. Para esto se necesita, tal como he argumentado, una infestación de memes, competencias cognitivas (hábitos, maneras) diseñadas en alguna otra parte y luego instaladas en nuestros cerebros, hábitos que cambian de raíz la arquitectura cognitiva de esos cerebros, convirtiéndolos en mentes. Hasta ahora, los únicos animales cuyos cerebros están pertrechados de este modo somos nosotros, Homo sapiens.

Igual que la célula eucariótica surgió en una ocasión comparativamente súbita de transferencia tecnológica, en que dos estirpes independientes de I+D se unieron en un único éxito de simbiosis para impulsarse en un gran salto adelante, la mente humana, la mente que comprende, es (y tuvo que ser) un producto de la simbiosis, uniendo los frutos de dos estirpes

básicamente independientes de I+D. He planteado que empezamos con

cerebros animales que, en gran parte, han sido rediseñados para ser una excelente base de recursos de pensamiento diseñados en otra parte: los memes, entre los que destacan las palabras. Nosotros adquirimos la mayoría de vocabulario inconscientemente, en el sentido siguiente: no nos dimos cuenta de estar aprendiendo siete palabras nuevas cada día cuando éramos niños y para la mayoría de palabras (palabras que no nos explican explícitamente) solo acotamos sus significados gradualmente, gracias a los procesos inconscientes que hallan patrones en nuestras experiencias previas con esas palabras. Una vez poseemos las palabras, podemos empezar a usarlas, pero no tenemos por qué darnos cuenta de lo que estamos haciendo (para cada palabra de vuestro vocabulario, hubo un caso debutante, la primera vez que la usasteis, en un acto de habla o en un monólogo interno o alguna cavilación; ¿cuántas veces habéis sido conscientes de hacerlo con las nuevas palabras que han entrado en

vuestro vocabulario habitual en los últimos años? ¿Alguna vez?). En cuanto las palabras se convierten en nuestros recursos habituales, y no en unos meros sonidos asociados a contextos, podemos empezar a usarlas para crear nuevos puntos de vista acerca de todo lo que nos encontramos.

Hasta el momento, hay pocos indicios de que esté surgiendo este tipo de fenómeno en las crecientes competencias de las máquinas de aprendizaje profundo. Tal como señala Domingos, las máquinas de aprendizaje están diseñadas (muy inteligentemente) para hacer uso de procesos de autorrediseño «darwiniescos» de abajo arriba. Para el Watson de IBM, el programa que en 2011 venció a los campeones del concurso televisivo Jeopardy!, Ken Jennings y Brad Rutter, las palabras con las cuales era competente uniéndolas para formar respuestas ganadoras no eran recursos de pensamiento, sino solo nodos situados en un espacio multidimensional de otros nodos, no tanto memes como restos fósiles de los memes humanos, que conservan ingentes cantidades de información sobre las creencias y prácticas humanas sin que sean participantes activos en tales prácticas. Aún no… pero quizá sí algún día. En pocas palabras, Watson aún no piensa pensamientos empleando las palabras acerca de las cuales tiene tanta información. Watson puede responder a preguntas (en realidad, gracias al curioso formato de Jeopardy! Watson puede formular preguntas de las cuales las pistas de Jeopardy! son las respuestas; pista de Jeopardy!: «La capital de Illinois»; concursante: «¿Qué es Springfield?»), pero esto no es entablar una conversación.

Lo que da a las mentes su facultad innovadora es la capacidad de autoobservarse, de someter los patrones de reacción del cerebro a una nueva ronda (o dos, o tres, o siete) de determinación de patrones.[2] En el ambiente actual del aprendizaje automático son los usuarios humanos, como Frances Arnold en su grupo de trabajo de proteínas y David Cope con sus experimentos en inteligencia musical, los diseñadores y otros operarios de las máquinas, los que están en posición de representar estos papeles, evaluando, ajustando, criticando, modificando y descartando los resultados sospechosos que surgen a menudo. Son los críticos cuyas actividades de control de calidad proporcionan las fuerzas selectivas adicionales que «en principio» podrían llevar estos sistemas hacia la comprensión, ascendiéndolos de meras herramientas a colegas, pero se trata de un paso gigante, o de una serie de pasos gigantes. Desde este punto de vista, podemos ver con más claridad que nuestras mentes infestadas de memes albergan a usuarios, críticos de los veredictos

brutos de nuestros cerebros animales, sin los cuales seríamos tan ingenuos como los otros mamíferos, astutos en su propio terreno pero incapaces ante cualquier novedad importante.

La curiosidad mató al gato, según un famoso meme inglés, y la curiosidad animal, impulsada de abajo arriba por la presencia de novedades, es un importante rasgo de alto riesgo pero gran recompensa en muchas especies, pero solo los seres humanos tenemos la capacidad de embarcarnos en una curiosidad controlada, sistemática, previsora y sometida a pruebas de hipótesis, una característica de los usuarios que aparecen en cada cerebro, usuarios que pueden aprovecharse de la enorme capacidad de los cerebros para descubrir regularidades estadísticas. La ilusión de usuario de la conciencia tiene el mismo papel en cada uno de nosotros que las interfaces entre humano y ordenador de Watson y otros sistemas de aprendizaje profundo; ofrecen algo así como un escaparate de talentos, un «mercado de ideas» en que la evaluación y la competición instantáneas pueden mejorar la velocidad y la resolución del control de calidad.

Así pues, el pensamiento humano (tal como identificó Darwin en el fenómeno que denominó selección metódica) puede acelerar la selección natural focalizando el poder selectivo mediante sistemas perceptuales y motivacionales de los domesticadores. Frances Arnold no está cultivando sus proteínas sin más; está enfrascada en un engendramiento intensivo y dirigido de las nuevas proteínas. Esto debería ponernos en estado de alerta ante la perspectiva de que nuestras maravillosas mentes no son inmunes a las modas y a los antojos que sesgan nuestros esfuerzos de autorrediseño de maneras sorprendentes e incluso autodestructoras. Al igual que los disparatados plumajes estimulados por los entusiastas y metódicos colombófilos o las lastimosas discapacidades pacientemente diseñadas en diversas variedades de perros miniatura, los seres humanos también pueden moldear sus mentes y convertirlas en artefactos grotescos (a menudo con ayuda de cómplices ansiosos) que las vuelven inútiles o aún peor.

Todo esto sugiere (aunque no demuestra) que sin que nosotros, los usuarios de las máquinas, interpretemos los resultados, con ojos críticos y perspicaces, los sistemas de aprendizaje profundo pueden aumentar sus competencias, sobrepasando a los cerebros animales (incluido el nuestro) en varios órdenes de magnitud para las tareas de abajo arriba de

búsqueda de regularidades estadísticas, pero nunca lograrán (nuestro tipo de) comprensión. Alguien puede objetar: «Muy bien, ¿y qué? El tipo de comprensión de abajo arriba de las computadoras acabará por engullir al tipo de comprensión humano, abrumándolo con el brutal tamaño y velocidad de su aprendizaje». El último avance en inteligencia artificial, AlphaGo, el programa de aprendizaje profundo que hace poco ha vencido a Lee Seedol, considerado por muchos el mejor jugador humano de go, respalda esta posibilidad, al menos en un aspecto, aunque no en otros. Ya he señalado que Frances Arnold y David Cope hacen un papel clave de control de calidad en los procesos de generación que controlan, como críticos cuyo juicio científico o estético decide qué caminos se explorarán con más detalle; podríamos decir que están guiando las máquinas de exploración que han diseñado a través del espacio de diseño. Pero AlphaGo hace algo parecido, según los trabajos que se han publicado: la forma de mejorar su juego es jugar miles de partidas de go contra sí mismo, introduciendo pequeñas mutaciones exploradoras en todas ellas, evaluando cuáles representan un probable progreso y usando tales evaluaciones para ajustar las futuras sesiones de partidas de entrenamiento. No es más que otro nivel de generación y prueba, en uno de los juegos más abstractos y aislados del ruido del mundo real y sus preocupaciones asociadas, pero AlphaGo aprende a hacer juicios «intuitivos» sobre situaciones con muy pocos puntos de referencia nítidos, esos puntos de referencia que los programas informáticos son tan duchos en revisar. Con los vehículos autónomos casi a punto para lanzarse al mercado (una perspectiva que no hace tantos años se consideraba excesivamente optimista y pocos tomaban en serio), ¿tardaremos mucho en tener una exploración científica también autónoma, «sin conductor»?

En definitiva, puede que pronto los juicios prácticos, estéticos y científicos se puedan descargar, se puedan «subcontratar», a agentes artificiales. Si Susan Blackmore está en lo cierto, esta abdicación o alienación del juicio humano ya está en marcha en el mundo digital de la música popular y los memes de Internet, o tremes, en su nueva terminología (véase capítulo 11). Durante siglos ha habido un exceso de memes, con quejas que se remontan a los primeros días de la imprenta, y desde entonces las personas no han dudado en pagar por disponer de filtros para quitarse de encima los más abrumadores, irritantes e inútiles, de un modo u otro. No intentéis leer todos los poemas de todos los poetas; esperad, en cambio, a que algún poeta o crítico acreditado edite

una buena antología. Pero ¿quién tiene la autoridad y acreditación suficiente? ¿Cuál coincidirá más con vuestros gustos y necesidades? Podéis suscribiros a alguna revista literaria que periódicamente reseña estas obras, junto con las de poetas individuales, pero ¿en cuál de ellas confiar? Quizá podéis informaros sobre su reputación y calidad en otras revistas. Uno puede ganarse la vida ocupándose de las necesidades expresadas por los demás buscadores de memes y si el negocio va de baja, podéis intentar crear nuevas necesidades, que podréis facilitar inmediatamente. Una situación muy familiar. Pero estamos entrando en una época en que los filtros, los creadores de modas y tendencias y los críticos puede que ya no sean personas, sino agentes artificiales. Esto es algo que no complacerá a todo el mundo, como veremos en el siguiente apartado; pero esta queja no impedirá la proliferación de capas y capas jerárquicas de estos replicadores diferenciales, con lo que puede que, al final, nos enfrentemos al desastre del aprendiz de mago y sus inacabables escobas.

En un anuncio de televisión de IBM, Watson, «conversando» con Bob Dylan dice que puede «leer 800 millones de páginas por segundo». Google Translate, otro icono de las máquinas de aprendizaje, ha barrido a los sistemas GOFAI y sus intentos de arriba abajo de analizar e interpretar (y, en consecuencia, comprender, en un pálido reflejo de la comprensión humana) el lenguaje humano; Google Translate es un traductor rápido y bueno (aunque aún lejos de la perfección), pero depende totalmente del corpus de traducciones ya efectuadas por seres humanos (y por informadores voluntarios bilingües, a quienes se les invita a colaborar en el sitio web). Buscando patrones y cribando millones de fragmentos que ya han sido correctamente traducidos (o lo suficiente como para encontrarse en línea), Google Translate se decide por una posible (probable) traducción aceptable sin tener la más mínima comprensión.

Es una afirmación polémica que requiere un análisis cuidadoso. Se cuenta una historia de un español que afirma: «Los franceses lo llaman couteau, los italianos, coltello, los alemanes Messer y los ingleses, knife; nosotros, en español, lo llamamos cuchillo, ¡y eso es lo que es!». El autocomplaciente aislamiento del español está asociado a algo que él conoce (lo que es un cuchillo), y esto parece que no tiene ningún equivalente en el «conocimiento» de Google Translate. En la jerga de la ciencia cognitiva, el conocimiento del español del significado de cuchillo

(y de couteau y de otros términos sinónimos) se basa en un conocimiento no lingüístico, una familiaridad con los cuchillos, con cortar y afilar, con el peso y el tacto de un trinchante, con el uso de un cortaplumas, etc. Ese español tiene, con respecto a la palabra cuchillo, lo que vosotros seguramente no tenéis respecto a la palabra tahalí, incluso aunque supierais que los ingleses le dicen baldric. Pero no vayamos tan rápido, porque sin duda Google Translate posee un abundante corpus de datos sobre los contextos en que aparece el término cuchillo, en que habrá palabras como corte, afilado, arma, llevar, manejar, estocada, tallar, cortar, tajar, comida, pan, filo, empuñadura, tahalí y muchas más, cada una con sus respectivas palabras asociadas. ¿No es toda esta información refinada y procesada sobre contextos lingüísticos una especie de conocimiento básico sobre la palabra cuchillo? Al fin y al cabo, ¿no es el mismo tipo de conocimiento básico que muchos de nosotros tenemos sobre términos técnicos como ARN mensajero o bosón de Higgs? Es una información que guía el proceso de traducción por canales cada vez más adecuados. Si confías en Google Translate para ser tu traductor bilingüe, rara vez te decepcionará. ¿No demuestra esto un considerable grado de comprensión? Muchos responderán «¡No!»; pero si queremos que este inflexible rechazo sea algo más que un mero desafío ritual de la máquina, mejor que haya algo que los seres comprendedores «de verdad» puedan hacer con su comprensión y que esté más allá de los poderes de Google Translate.

Probemos con esto: una cosa es traducir un examen final del inglés al castellano y otra cosa es evaluarlo. Pues resulta que esto no nos servirá, porque el innovador trabajo de Thomas Landauer sobre el «análisis semántico latente» (véase, p. ej., Littman et al. 1998) ya ha creado un programa informático que hace precisamente esto (Rehder et al. 1998). Un profesor plantea una cuestión a desarrollar en un examen y redacta una respuesta de sobresaliente a esa cuestión, que a continuación se facilita a un programa informático y a un becario humano como ejemplo de lo que sería un buen desarrollo de esa cuestión (por supuesto, la respuesta de sobresaliente no se muestra a los examinandos). A continuación, tanto el programa como el becario evalúan todas las respuestas de los estudiantes; el resultado es que las notas que pone el programa se acercan más a las del profesor que las del becario, que se supone que es un experto en el campo. Es un resultado, como mínimo, inquietante; he aquí un programa informático que no entiende inglés, y mucho menos el tema del examen, pero al cual le bastan las sofisticadas

propiedades estadísticas de las respuestas modelo del profesor para evaluar las respuestas de los estudiantes a esas mismas preguntas con una alta fiabilidad. ¡Competencia para evaluar sin pizca de comprensión! (Landauer acepta que, en principio, un estudiante podría idear un desarrollo a la cuestión que fuera un sinsentido absoluto pero que tuviera las propiedades estadísticas correctas; ahora bien, un estudiante que pudiera hacer esto bien merecería un sobresaliente).

Probemos, pues, con otra cosa: ¿qué tal la tarea de tener una conversación sensata con un ser humano? Esto no es nada más que el clásico test de Turing, que puede separar el grano de la paja de modo muy definitivo. Watson puede vencer a Ken Jennings y a Brad Rutter, dos campeones humanos del concurso Jeopardy!, pero no se trata de una conversación libre, y los anuncios en que Watson charla con Jennings o Dylan o una joven superviviente de cáncer (interpretada por una actriz) son guiones, no conversaciones improvisadas. Una conversación real y abierta entre dos agentes hablantes es, tal como observó Descartes (1637) en su clarividente suposición de un autómata hablante, una muestra espectacular de grandes habilidades cognitivas (si no infinitas, tal como Descartes aventuró). ¿Por qué? ¡Porque una conversación humana normal tiene lugar en un espacio de posibilidades regido por motivaciones flotantes griceanas! Puede que yo no tenga la intención expresa de que tú identifiques mi intención de que tú creas que lo que digo es cierto (o que es ironía, o una broma, o una exageración evidente), pero si no eres capaz de efectuar este tipo de identificación, y si tampoco eres capaz de generar actos de habla con motivaciones flotantes parecidas que expliquen tus propias respuestas y retos, no serás un conversador convincente y cautivador. Las sucesivas capas de cognición de Grice quizá no representan con exactitud los rasgos inmediatos de una representación, pero sí que especifican una competencia.

Un participante en una conversación de alto nivel tiene que poder reconocer patrones en sus propias acciones y reacciones verbales, plantear situaciones hipotéticas, captar los chistes, identificar los faroleos, cambiar de tema cuando la cosa se vuelve aburrida, explicar sus actos previos de habla cuando se le pide, etc. Todo esto requiere (si no es magia) la representación de todas las distinciones que, de algún modo, tienen que identificarse para establecer el escenario de las acciones emprendidas, mentales y, finalmente, verbales. Por ejemplo, si no os dais cuenta o no podéis daros cuenta (en algún sentido básico, quizá

subliminal) de que estoy bromeando, no podréis seguir la corriente, excepto por accidente. Esta identificación, este «darse cuenta de», no es solo cuestión de que vuestro cerebro animal haga una distinción; más bien se trata de algún tipo de influencia agudizada que no solo distingue retrospectivamente lo que identifica de sus competidores en ese momento, sino que también, y con la misma importancia, contribuye a la creación de un identificador duradero, un «ejecutivo» relativamente duradero que no es un lugar en el cerebro, sino una especie de coalición política que puede asumir el control de las competiciones subsiguientes durante un cierto tiempo. Tales diferencias de resultado («y ahora, ¿qué?») pueden ser sorprendentes.

Imaginemos que nos piden que completemos palabras parciales (la «tarea de compleción de raíces») y nos presentan:

gra ___

o

ven ___

¿Qué se os ha ocurrido? ¿Habéis pensado en grano, grado o grapa, por ejemplo, y venta, venda o vencejo? Supongamos que, pocos segundos después de aparecer la palabra a completar, en la pantalla se muestra, solo muy brevemente, esto:

grano

y luego, un segundo más tarde, nuevamente gra ___. La tentación de responder grano sería enorme, sin duda. Pero supongamos que los

experimentadores os han dicho: «Si veis aparecer una palabra en la pantalla, no la uséis como respuesta»; en tal caso, podréis superar las ansias de decir grano y, en general, diréis algo diferente, como gracia, granizo o granada. Es poco probable que digáis grano, porque podréis respetar la prohibición que os ha planteado el experimentador. Pero esto solo sucede si os dais cuenta (o sois conscientes) de que ha aparecido brevemente la palabra prohibida. Si la palabra solo aparece durante 50 milisegundos y luego aparece una «máscara» (una pantalla con un patrón) durante 500 milisegundos, aumenta la probabilidad de que digáis grano, a pesar de intentar respetar la prohibición (Debner y Jacoby 1994). Fijémonos en la elegancia de este diseño experimental: hay dos grupos de participantes, a uno se le dice que use la palabra «implícita», la palabra «primada», en la jerga de psicología, si es una buena respuesta, al otro se le dice que no use la palabra «implícita» si es una buena respuesta. Ambos grupos reciben esta «imprimación» durante 50 ms o 500 ms, seguida de una máscara; esta máscara no enmascara las imprimaciones de 500 ms (los participantes las identifican, pueden decir cuáles son, son conscientes de ellas y pueden usarlas o contenerse de usarlas, según lo ordenado), pero sí las de 50 ms, pues los participantes aseguran no haber visto ninguna palabra (es un fenómeno habitual de «enmascaramiento a posteriori»). En ambos casos, el cerebro efectúa una identificación de la imprimación, tal como demuestra que, en el caso de la prohibición, las imprimaciones breves aumentan la probabilidad de que esa sea la respuesta, mientras que las imprimaciones largas la disminuyen. Dehaene y Naccache (2001) remarcan la «imposibilidad de que los sujetos [es decir, ejecutivos] usen estratégicamente la información inconsciente».

Mi afirmación, pues, es que el aprendizaje profundo (hasta el momento) identifica y distingue, pero no se da cuenta. Es decir, el caudal de datos que acepta un sistema no tiene relevancia para el sistema, excepto como más comida a digerir. Postrado en cama, sin tener que valerse por sí mismo, no tiene más objetivos que aumentar su reserva de información bien ordenada. Aparte de la capacidad que compartimos con Watson y otros sistemas de aprendizaje profundo de adquirir conocimientos prácticos que dependen de regularidades estadísticas que obtenemos de la experiencia, hay la capacidad de decidir qué buscar y por qué, a partir de los objetivos propios actuales. Es la ausencia de razones prácticas, de inteligencia empleada para emprender finalidades autogeneradas, diversas y cambiantes, lo que (actualmente) diferencia a Watson de

cualquier persona normal sana. Cuando Watson llegue (si llega) al grado de sofisticación en que pueda enfrascarse en la práctica humana de aportar y evaluar razones, dejará de ser una mera herramienta y se convertirá en un colega; en este punto es cuando Watson, y no solo sus creadores y conservadores, empezará a ser considerado responsable de sus actos.

La forma en que las máquinas de aprendizaje profundo dependen de la comprensión humana merece un análisis más detallado. En el capítulo 8 consideramos la atrevida crítica de Deacon de la IA tradicional: los potenciales diseñadores de mentes que abstraen los requisitos de captura de energía y de autodefensa, limitando así su búsqueda a sistemas parasitarios, siempre dependientes del mantenimiento humano: son herramientas, no colegas. Ahora podemos ver que el tipo de comprensión que los sistemas de IA muestran actualmente (y que es cada vez más comparable a la mejor de las comprensiones humanas) también es parasitaria, con una dependencia radical del enorme legado de comprensión humana que puede aprovechar. Google Translate no sería nada sin los millones de buenas traducciones efectuadas por humanos en las que se basa, y el descomunal compendio de conocimientos factuales de Watson depende igualmente de los millones de páginas que absorbe cada día de Internet. Podemos adaptar la famosa frase de Newton y decir que estos programas se alzan sobre los hombros de gigantes, sirviéndose de toda la inteligencia que se muestra en los productos previos del diseño inteligente.

Todo esto queda bien ilustrado con un problema que planteé a mis estudiantes cuando Watson venció a Jennings y Rutter en Jeopardy! Les puse unos deberes consistentes en buscar cuestiones que bloquearan a Watson pero que fueran fáciles para Jennings o Rutter (o para cualquier otro ser humano normal). Cabe señalar que, en Jeopardy!, se tuvieron que ajustar las reglas a favor de Watson; por ejemplo, los problemas planteados eran todos verbales, sin implicar nunca la visión ni la audición. En mi opinión, las probables cuestiones difíciles o imposibles para Watson implican siempre algún tipo de imaginación:

La juguetona palabra que puedes hacer con dos rosquillas y dos tirachinas.

Respuesta: ¿Qué es un yoyó?

Un cambio de una letra que convierte un árbol en un mamífero.

Respuesta: ¿Qué es un alce respecto a un arce?

El número entre 0 y 9 que haría unas buenas gafas.

Respuesta: ¿Qué es el ocho?

Tengo mejores ejemplos, pero no los publicaré ni los pondré en Internet, para que Watson no los consiga y los tenga a punto para otro concurso futuro. Watson no necesita imaginación cuando puede pescar en la imaginación de los demás. Fijaos que, en este aspecto, Watson es profundamente darwiniano: ni él ni la selección natural dependen de la previsión ni de la imaginación, porque están dirigidos por procesos incansables y sin comprensión que extraen información (patrones estadísticos que pueden orientar mejoras de diseños) a partir de lo que ya ha pasado. Ambos son ciegos para todo aquello que no ha sucedido al alcance de sus procesos de selección. Evidentemente, si no hay nada nuevo bajo el Sol, esto no representa limitación alguna, pero la imaginación humana, la capacidad que tenemos de imaginar realidades que no nos son accesibles sin más que encaramarnos a una colina, parece que es una jugada clave, que nos permite crear, mediante un diseño previsor, oportunidades y, finalmente, iniciativas y artefactos que no podrían surgir de otro modo. Una mente humana consciente no es ningún milagro, no es ninguna violación de los principios de la selección natural, sino una novedosa ampliación de estos, una nueva grúa que perfecciona el concepto de posibles adyacentes del biólogo evolutivo Stuart Kauffman: nosotros tenemos muchos más lugares adyacentes en el espacio de diseño porque hemos evolucionado la capacidad de pensar acerca de ellos y o bien buscarlos o bien evitarlos. La cuestión pendiente para Domingos y otros representantes del aprendizaje profundo es si aprender una teoría lo bastante detallada y dinámica de agentes con imaginación y capacidades de ofrecer razones permitiría a un sistema (un

programa informático, un Algoritmo Maestro) generar y explotar las capacidades de tales agentes; es decir, generar todos los poderes moralmente relevantes de una persona.[3]

Mi punto de vista sigue siendo que el aprendizaje profundo no nos dará (en los próximos cincuenta años) nada parecido a la «inteligencia sobrehumana» que tanta atención y alarma ha suscitado últimamente (Bostrom 2014; referencias anteriores son Moravec 1988; Kurzweil 2005 y Chalmers 2010; véase también la pregunta anual de 2015 de la Edge Foundation y Katchadourian 2015). El acelerado crecimiento de las competencias en la IA, bajo la égida del aprendizaje profundo, ha cogido por sorpresa también a muchos profesionales de la disciplina, no solo a los comentaristas y críticos, como yo mismo. Hay una larga tradición de autobombo y exageración en la IA, que se remonta a sus primeros días, lo que ha hecho que muchos de nosotros descartemos siempre un 70% o más, pongamos por caso, de los últimos «avances revolucionarios», pero cuando grandes expertos tecnológicos como Elon Musk y grandes científicos como Martin Rees y Stephen Hawking empiezan a hacer llamadas de alarma acerca de la posibilidad de que la IA nos lleve pronto a una desaparición cataclísmica de la civilización humana, es el momento de refrenar los hábitos y revisar las sospechas propias. Y tras hacerlo, mi veredicto no ha cambiado, pero sí es más conjetural de lo que solía ser. Siempre he afirmado que la «IA fuerte» es «posible en principio», pero lo consideraba una posibilidad práctica insignificante, porque resultaría demasiado costosa y no nos aportaría nada que realmente necesitáramos. Domingos y otros me han demostrado que puede haber vías practicables (tanto en los aspectos técnicos como económicos) que había subestimado, pero sigo pensando que la tarea está varios órdenes de magnitud, en dimensiones y en dificultad, por encima de lo que afirman sus seguidores, por las razones que he presentado en este capítulo y en el capítulo 8 (el ejemplo del Nuevayorkbot).

En resumidas cuentas, no me preocupa que la humanidad vaya a crear una raza de agentes superinteligentes destinados a esclavizarnos, pero eso no significa que no me preocupe en absoluto. Creo que en nuestro futuro inmediato hay otras situaciones, menos alarmantes pero más probables, que deben preocuparnos y obligarnos a emprender acciones inmediatas.

¿QUÉ SERÁ DE NOSOTROS?

Ya existen artefactos (y muchos más en desarrollo) con competencias tan superiores a cualquier competencia humana que se apropiarán de nuestra autoridad como expertos, una autoridad que no se ha puesto nunca en cuestión desde el inicio de la era del diseño inteligente. Y cuando cedamos la hegemonía a estos artefactos, será por buenas razones, prácticas y morales. Actualmente sería de una negligencia criminal embarcar a pasajeros en un crucero sin equipar el barco con varios sistemas GPS; la navegación celeste con sextante, brújula, cronómetro y almanaques náuticos es un vestigio pintoresco de competencias obsoletas, como lo es afilar una guadaña o arar con mulas. Los que nos deleitamos con estas reliquias del pasado, guadañas en el campo o sextantes en el mar, podemos embarcarnos con nuestros artilugios pasados de moda y practicar con ellos, en el caso improbable de que necesitemos un sistema de respaldo, pero no tenemos ningún derecho a arriesgar las vidas de los demás evitando los aparatos de alta tecnología disponibles.

Todos aprendemos aún las tablas de multiplicación hasta la del 10, y luego aprendemos cómo usarlas para valores aún más altos (sí, ¿verdad?), y podemos hacer divisiones con lápiz y papel, pero pocos sabemos o recordamos el algoritmo para extraer raíces cuadradas. ¿Y qué? No es necesario gastar esfuerzos y células del cerebro en tareas que se pueden ejecutar sin más que pulsar unas pocas teclas o preguntando a Google o Siri. La respuesta habitual de los aprensivos es que al educar a nuestros hijos tenemos que enseñarles los principios básicos de todos los métodos en los cuales estamos versados, y para esta comprensión resulta valioso un cierto grado de experiencia real con tales métodos, pero es posible que podamos afianzar esos principios básicos sin someter a nuestros hijos a estas tareas monótonas y pasadas de moda, ¿no es cierto? Esto parece sensato, pero ¿podemos generalizarlo?

Pensemos en la formación médica. Watson no es más que uno de muchos sistemas informáticos que empiezan a superar a los mejores diagnosticadores y especialistas en su propio terreno. ¿Estaríais dispuestos a consentir que vuestro doctor continuara siendo un lector «intuitivo» de síntomas a la vieja usanza en lugar de confiar en un

sistema informático que ha demostrado ser cien veces más fiable que cualquier especialista a la hora de establecer diagnósticos poco comunes? Vuestro seguro médico os obligaría a aceptar las pruebas y los médicos concienzudos verían que tienen que sofocar sus anhelos de ser unos héroes del diagnóstico y rendirse a la mayor autoridad de las máquinas cuyos botones pulsan. ¿Qué implica todo esto para la formación de futuros médicos? ¿Se alentará a que dejen de lado partes de la educación médica tradicional, como la anatomía, la fisiología o la bioquímica, junto con la capacidad de hacer divisiones a mano o de leer un mapa? «Lo tomas o lo dejas» es el lema que tenemos que citar en este momento, y tiene muchos ejemplos positivos. ¿Pueden vuestros hijos leer mapas de carreteras tan fácilmente como vosotros o se han vuelto dependientes del GPS para encontrar el camino? ¿Debemos preocuparnos por atontarnos a causa de nuestra creciente confianza en máquinas inteligentes?

Hasta ahora, hay una frontera bien definida entre máquinas que mejoran nuestras capacidades intelectuales «periféricas» (de percepción, de cálculo algorítmico y de memoria) y las máquinas que como mínimo pretenden sustituir nuestras capacidades intelectuales «centrales» de comprensión (incluida la imaginación), planificación y toma de decisiones. Las calculadoras de bolsillo, los sistemas GPS, los sistemas gráficos de Pixar para interpolar fotogramas, calcular sombras y ajustar texturas y la PCR y la CRISPR en genética son instrumentos y técnicas en el lado periférico de la frontera, aunque realicen tareas que no hace mucho requerían una pericia considerable. Es de esperar que la frontera se reduzca, volviendo rutinarias cada vez más tareas cognitivas, lo que está muy bien siempre y cuando tengamos claro dónde está la frontera en cada momento. En mi opinión, el verdadero peligro no es que máquinas más inteligentes que nosotros usurpen nuestro papel como comandantes de nuestros destinos, sino que sobrestimemos la comprensión de nuestros más recientes recursos de pensamiento, cediéndoles prematuramente una autoridad por encima de sus competencias reales.

Siempre hay formas de reforzar la frontera mientras se va reduciendo, haciéndola bien visible para todo el mundo. Sin duda, habrá innovaciones que traspasarán esta línea y, si la historia reciente nos puede servir de guía, seguro que cada nuevo avance se sobrevalorará. Existen antídotos, que deberíamos esforzarnos en obtener. Sabemos que las personas adoptan rápidamente el enfoque intencional ante cualquier

cosa que les impacta como inteligente, y dado que la suposición predeterminada del enfoque intencional es la racionalidad (o la comprensión), hay que tomar medidas para mostrar a las personas cómo moderar su credulidad al interaccionar con un sistema antropomórfico. En primer lugar, deberíamos revelar y ridiculizar cualquier tipo de antropomorfismo gratuito en los sistemas, las voces agradables y cada vez más humanas, las animadas acotaciones (por muy enlatadas que sean). Al interactuar con un ordenador, deberíamos saber que estamos interactuando con un ordenador. Los sistemas que ocultan deliberadamente sus atajos y sus carencias deberían considerarse fraudulentos y sus creadores deberían ir a prisión por el crimen de crear y usar una inteligencia artificial que se hace pasar por un ser humano.

Deberíamos fomentar el desarrollo de una tradición de hipermodestia, acompañando siempre toda la propaganda con una lista obligatoria de los límites conocidos, las carencias, las lagunas y demás fuentes de ilusión cognitiva (de la misma forma en que obligamos a las empresas farmacéuticas a recitar las largas listas de efectos secundarios siempre que anuncian un nuevo fármaco). Concursos para revelar los límites de comprensión, al estilo del test de Turing, serían una buena innovación, que alentaría a la gente a valorar su capacidad de calar las máquinas fraudulentas de la misma manera en que se enorgullecen de identificar a un farsante. ¿Quién puede hallar la forma más rápida y segura de revelar los límites de este aparato inteligente? (no deja de ser irónico que la tolerancia y la buena educación que fomentamos en nuestros hijos al tratar con desconocidos tiene el efecto indeseado de convertirlos en crédulos usuarios de los diversos no-agentes verbales con los que se encuentran; deben aprender a ser agresivos, inquisitivos e impertinentes al tratar con estos nuevos «ayudantes»).

Es de esperar que las nuevas prótesis cognitivas seguirán diseñándose para ser parásitas, para ser herramientas y no colegas. Su único objetivo «innato», determinado por sus creadores, debería ser responder de un modo constructivo y transparente a las demandas del usuario. Un motivo de preocupación es que las máquinas de aprendizaje serán cada vez más competentes en descubrir lo que queremos nosotros, sus usuarios, y, en consecuencia, pueden diseñarse para ocultarnos sus extrapolaciones «útiles». Ya somos bien conscientes de la frustración de las «correcciones» no deseadas de los correctores ortográficos, y por eso muchos desactivamos estas funciones, pues su capacidad de interpretar

erróneamente nuestras intenciones aún es demasiado alta en muchos casos. Estamos aquí ante la primera capa de semicomprensión con la que tenemos que tratar.

Ya van surgiendo algunas tensiones alrededor de los desarrollos actuales, que vale la pena comentar. Google tiene un programa para mejorar su motor de búsqueda descifrando automáticamente lo que sospecha que el usuario quiere decir realmente cuando introduce su cadena de caracteres (http://googleblog.blogspot.com/2010/01/helping-computers-understand-language.html); no hay duda de que esto sería útil para muchos propósitos, pero no para todos. Tal como ha remarcado Douglas Hofstadter en una carta abierta a un antiguo estudiante, en ese momento trabajando en Google en este proyecto:

Me preocupa y, de hecho, me molesta soberanamente, que Google intente socavar continuamente las cosas de las que dependo a diario.

Cuando en una búsqueda de Google pongo algo entre comillas, quiero que se tome de modo literal, y con razón. Por ejemplo (y este no es más que un ejemplo entre muchos), como escritor minucioso, siempre estoy intentando determinar la mejor manera de expresar algo en un idioma u otro, de modo que muchas veces compruebo dos expresiones posibles y las comparo para ver si una de ellas aparece con mucha más frecuencia que la otra. Para mí, es una forma muy importante de descubrir cosas acerca de diversas expresiones, pero si Google no las toma de modo literal y no duda en cambiar caprichosamente algunas palabras por otras en lo que acabo de escribir, me estará engañando y dando una información incorrecta cuando yo veo que una determinada frase aparece con mucha frecuencia. Todo esto me resulta muy molesto.

Quiero que las máquinas sean fiables y mecánicas, no que estén siempre desviándose de lo que yo les pido que hagan. La presunta «inteligencia» de las máquinas puede ser útil de vez en cuando, pero también puede resultar muy contraproducente e incluso perjudicial y, en mi experiencia, la inteligencia artificial (y aquí uso artificial en el sentido de falso, ilegítimo) que actualmente se instala en todos los dispositivos tecnológicos siempre me resulta muy decepcionante.

Es por ello por lo que no me entusiasma lo que tu grupo está haciendo; en realidad, me preocupa mucho. Es solo otro intento de hacer dispositivos mecánicos que no sean fiables como tales. Le pides a Google que haga X, suponiendo que hará precisamente X, pero en realidad hace Y, y este Y es lo que él «piensa» que tú querías decir. En mi opinión, este intento de leer mi mente es increíblemente molesto, cuando no peligroso, porque casi nunca es correcto, muchas veces ni por asomo. Quiero máquinas fiables y mecánicas, de modo que sepa con seguridad con lo que estoy tratando. No quiero que sean más astutas que yo, porque, en definitiva, lo que harán me engañará y me confundirá. Diría que esto es un punto muy básico, pero parece que Google lo ignora por completo (o, por lo menos, en tu grupo). Creo que es un tremendo error. (Correspondencia personal con Abhijit Mahabal, 2010)

Como poco, estos sistemas deberían: 1) anunciar con claridad cuándo intentan «leer la mente» y no ser meramente «mecánicos»; y 2) ofrecer a los usuarios la opción de desactivar la «comprensión» no deseada de la misma forma en que se puede desactivar un verificador ortográfico «con demasiada iniciativa». Una ley de «responsabilidad estricta» podría ofrecer el necesario incentivo de diseño: cualquiera que use un sistema de IA para tomar decisiones que impacten en las vidas y el bienestar de las personas, igual que los usuarios de otros equipos potentes y peligrosos, tiene que estar adecuadamente formado (y garantizado, quizá) y someterse a estrictas normas de responsabilidad, de modo que sea de su interés comportarse de modo escéptico, escrupuloso e inquisitivo en sus interacciones, por temor de dejarse engañar por sus propios dispositivos. Esto alentaría a los diseñadores de tales sistemas a hacerlos transparentes y modestos, ya que los usuarios rechazarían todos aquellos que les pudieran conducir a demandas de mala praxis.

Hay otra política que puede ayudar a mantener a raya la rendición de nuestras responsabilidades cognitivas. Pensemos en la tecnología para «hacernos más fuertes»: por un lado, tenemos la vía de la excavadora y por otra, la de una máquina de fitness. La primera te permite hacer proezas increíbles sin dejar de ser un debilucho de 45 kg; la segunda te permite obtener la fuerza necesaria para hacer grandes cosas por tu propia cuenta. La mayoría de software que ha potenciado nuestros poderes cognitivos ha sido del tipo excavadora, desde los telescopios y

microscopios a los secuenciadores y los nuevos productos del aprendizaje profundo. ¿Podría haber también software al estilo de las máquinas de fitness para desarrollar los «músculos» de la comprensión de los individuos? Pues sí, y de hecho, en 1985, George Smith y yo mismo, junto con los programadores Steve Barney y Steve Cohen, organizamos el Curricular Software Studio en Tufts, con el objetivo de crear «prótesis de imaginación», softwa-

re que equiparía y educaría las mentes de los estudiantes, desencallando cuellos de botella pedagógicos y permitiendo que elaborasen en su imaginación modelos armoniosos, dinámicos y precisos de fenómenos complejos, como la genética de poblaciones, la estratigrafía (la interpretación de la historia geológica de las capas de rocas), la estadística y la informática. El objetivo era idear sistemas que, una vez dominados, se pudieran dejar de lado porque los usuarios habían internalizado sus principios básicos y habían alcanzado el grado de comprensión que se obtiene con una exploración intensiva. Ahora quizá es el momento de proyectos mucho más ambiciosos diseñados para ayudar a las personas a pensar con creatividad y precisión acerca de los muchos fenómenos complejos con que nos topamos, de modo que puedan ser usuarios independientes, inteligentes y con comprensión de las prótesis epistemológicas en desarrollo, no solo unos beneficiarios pasivos y acríticos de los diversos cachivaches tecnológicos que reciben.

Hemos echado un vistazo a algunas de las innovaciones que nos han llevado a renunciar al dominio de la creación que desde hace mucho ha sido el sello distintivo de la comprensión en nuestra especie… y más que llegarán. Durante milenios, nuestra motivación ha sido la idea expresada en la frase de Feynman: «No entiendo lo que no puedo crear». Pero desde hace poco, nuestro ingenio ha dado lugar a un terreno resbaladizo: de forma indirecta, estamos construyendo cosas que solo entendemos en parte, cosas que, a su vez, pueden crear cosas que no entendemos en absoluto. Como algunas de estas cosas tienen unas capacidades asombrosas, podemos empezar a dudar del valor (o por lo menos del valor privilegiado) de la comprensión. «¡La comprensión está tan pasada de moda, es tan de otra época! ¿Quién quiere comprensión cuando todos podemos ser los beneficiarios de artefactos que nos ahorran todo este arduo trabajo?»

¿Tenemos alguna respuesta para esto? Necesitamos algo más que la tradición si queremos defender la idea de que la comprensión o bien es intrínsecamente buena (un bien en sí mismo, con independencia de todos los beneficios que proporciona indirectamente) o bien es una necesidad práctica si queremos seguir viviendo el tipo de vida que nos importa. Se espera que los filósofos, como yo, retrocedan con consternación ante un futuro así. Tal como dijo Sócrates «una vida que no se cuestiona, no merece la pena ser vivida», y desde Sócrates hemos dado por sentado que lograr una comprensión cada vez mayor de todo es nuestro más elevado objetivo profesional, si no nuestro más elevado objetivo en general. Pero tal como otro filósofo, Kurt Baier, añadió una vez, «una vida excesivamente cuestionada, tampoco es nada del otro mundo». La mayoría de personas se conforman con ser las beneficiarias de la medicina, la investigación científica y la creación artística sin tener mucha idea sobre cómo se ha creado toda esta «magia». ¿Tan terrible sería aceptar una vida hipercivilizada y dejar que nuestros artefactos fueran unos buenos administradores y cuidadores de nuestro bienestar?

No he podido idear ningún argumento convincente para la atractiva conclusión de que la comprensión es «intrínsecamente» valiosa (aunque creo que la comprensión es una de las mayores emociones de la vida), pero se puede plantear una buena argumentación para la necesidad de conservar y mejorar la comprensión humana y para protegerla de las variantes artificiosas de la comprensión que se están desarrollando con el aprendizaje profundo, por razones profundamente prácticas. Los artefactos se pueden estropear y si pocas personas los comprenden con el detalle suficiente para repararlos o sustituirlos por otros medios de realizar sus tareas, podemos encontrarnos, a nosotros y a todo aquello que nos importa, en una situación difícil. Muchas personas han señalado que, para algunos de nuestros artilugios de alta tecnología, hay cada vez menos técnicos capaces de repararlos, o incluso no hay ninguno. Comprar una nueva impresora multifunción resulta más barato que reparar la impresora averiada: tírala y compra una nueva. Los sistemas operativos de las computadoras siguen una versión parecida de esta misma política: cuando el software falla o se corrompe, no te molestes en intentar hallar y corregir el error, en deshacer la mutación que se ha introducido en el sistema; reinicializa todo y las últimas y límpidas versiones de tus programas preferidos, almacenadas a buen recaudo en la memoria, sustituirán a las copias defectuosas. Ahora bien, este proceso ¿hasta dónde puede llegar?

Consideremos una situación habitual de confianza sin comprensión en la tecnología. Un buen coche puede ser una delicia; te permite ir adonde necesitas ir, a tiempo, con fiabilidad y, en general, con una gran comodidad, con música, aire acondicionado y con el GPS guiándote. En los países desarrollados solemos dar por hecha la presencia de los coches, como un elemento más de la vida cotidiana, un recurso siempre disponible; planificamos nuestros proyectos teniendo en cuenta que dispondremos de algún coche. Pero cuando el coche se estropea, la vida se nos complica de golpe. A no ser que seamos un gran aficionado a la mecánica del automóvil, tendremos que confiar en toda una red de servicio de grúas, mecánicos, agentes de seguros, etc. En algún punto, decidimos dejar de lado nuestro viejo y achacoso coche y empezar con un nuevo modelo; y la vida sigue, sin apenas interrupción.

Pero pensemos en el sistema que hace posible todo esto: autopistas, refinerías de petróleo, fabricantes de automóviles, compañías de seguros, bancos, la bolsa, el gobierno. Nuestra civilización ha funcionado sin demasiados problemas (con algunas alteraciones graves) durante miles de años, creciendo en complejidad y poder. ¿Puede llegar a averiarse? Sí, sin duda, y entonces ¿a quién recurriremos para reparar el desaguisado? No podemos comprar una nueva civilización si la nuestra se colapsa, de modo que es mejor que mantengamos la que tenemos en buen funcionamiento. Pero ¿quiénes son sus mecánicos, sus técnicos de mantenimiento? Los políticos, los jueces, los banqueros, los empresarios, los periodistas, los profesores (en resumen, los líderes de la sociedad) se pueden equiparar más a un conductor mediocre que lo que nos gustaría pensar: se encargan de las pequeñas cosas que tienen asignadas para orientar su parte de todo el aparato, pero son felizmente ignorantes de las complejidades de las que depende el sistema completo. Según el economista y pensador evolutivo Paul Seabright (2010), la optimista estrechez de miras con la que operan no es un fallo lamentable y corregible del sistema, sino una situación posibilitadora. Esta distribución de la comprensión parcial no es optativa. Los edificios de la construcción social que conforman nuestras vidas en tantos aspectos dependen de nuestra miope confianza en la solidez de su estructura, que no necesita de nuestra atención.

En un fragmento, Seabright compara nuestra civilización con un castillo de termitas. Ambos son creaciones, maravillas de diseño ingenioso basado, a su vez, en más diseño ingenioso que se levanta sobre el terreno

circundante, una obra, en definitiva, de muchos individuos que actúan de común acuerdo. Ambos son subproductos de los procesos evolutivos que crearon y moldearon a esos individuos y, en ambos casos, las innovaciones de diseño que explican la remarcable resistencia y eficiencia observadas no son la creación de individuos concretos, sino el afortunado resultado de empresas miopes e inconscientes de esos individuos, a lo largo de varias generaciones. Pero también hay diferencias significativas. La cooperación humana es un fenómeno precario y notable, muy diferente de la cooperación casi ciega de las termitas y, de hecho, algo nunca visto en el mundo natural, un rasgo único con un linaje único en la evolución. Tal como hemos visto, depende de nuestra capacidad de involucrarnos unos con otros en el «espacio de razones», tal como lo expresa Wilfrid Sellars. Seabright argumenta que la cooperación depende de la confianza, una especie de pegamento social invisible que permite tanto proyectos grandiosos como terroríficos, y esta confianza no es un «instinto natural» preprogramado por la evolución en nuestros cerebros; es demasiado reciente.[4] La confianza es un subproducto de situaciones sociales que son, a la vez, su condición facilitadora y su producto más importante. Nos hemos impulsado a nosotros mismos hacia las embriagadoras alturas de la civilización moderna, donde nuestras emociones naturales y otras respuestas instintivas no siempre son adecuadas para las nuevas situaciones.

La civilización es una tarea en curso, y renunciamos a nuestro intento de comprenderla por nuestra cuenta y riesgo. Pensemos otra vez en el castillo de las termitas. Nosotros, observadores humanos, podemos apreciar su excelencia y su complejidad de modos que están fuera del alcance de los sistemas nerviosos de sus habitantes. También podemos aspirar a tener una perspectiva así de olímpica sobre muestro propio mundo artefactual, un logro al que solo los humanos podemos aspirar. Si no tenemos éxito, nos arriesgamos a desmantelar nuestras valiosas creaciones a pesar de nuestras mejores intenciones. La evolución en dos ámbitos, el genético y el cultural, nos ha otorgado la capacidad de conocernos a nosotros mismos. Pero a pesar de varios milenios de diseño inteligente cada vez más generalizado, aún estamos poco más que manteniéndonos a flote en un océano de misterios y problemas, muchos de los cuales creados por nuestros esfuerzos de comprensión; además, nos acechan peligros que pueden interrumpir nuestra búsqueda antes de

que nosotros o nuestros descendientes puedan satisfacer nuestra insaciable curiosidad.

POR FIN EN CASA

Con esto hemos completado nuestro viaje de las bacterias a Bach y otra vez de vuelta. Ha sido una excursión larga y dificultosa, a través de terrenos peligrosos, por regiones apenas transitadas por los filósofos y otras regiones asediadas por los filósofos y evitadas por los científicos. Os he invitado a cargar sobre los hombros algunas ideas claramente antiintuitivas y os he intentado mostrar cómo nos iluminan el camino. Ahora me gustaría presentar un resumen de los hitos más destacados que hemos superado y recordaros por qué los he considerado unas paradas necesarias en la ruta.

Empezamos con el problema de la mente y la radical polarización de Descartes del tema. Por un lado, las ciencias de la materia, el movimiento y la energía y cómo respaldan la vida, gracias a la evolución; por el otro, los fenómenos de la conciencia, íntimamente familiares pero, a la vez, misteriosos y privados. ¿Cómo se puede soldar esta fractura dualista? He planteado que el primer paso para solucionar el problema es la extraña inversión del razonamiento de Darwin, la idea revolucionaria de que todo el diseño que observamos en la biosfera puede ser, y en último término tiene que ser, el producto de procesos de selección natural ciegos, sin comprensión y sin propósito alguno. Ya no tenemos que ver a la mente como la causa de todo lo demás.

La evolución por selección natural puede revelar ciegamente las razones sin razonadores, las motivaciones flotantes que explican por qué las partes de los seres vivos están dispuestas como lo están, respondiendo a las dos preguntas ¿cómo es que? y ¿para qué? Darwin ofreció el primer gran ejemplo de competencia sin comprensión en el proceso de la selección natural. Más tarde, la extraña inversión del razonamiento de Turing proporcionó un ejemplo, y un banco de trabajo para explorar las posibilidades, de otra variedad de competencia sin comprensión: las computadoras, que a diferencia de los agentes humanos de los cuales

recibieron el nombre, no tienen que entender las técnicas que aprovechan con tanta competencia. Se pueden lograr tantas cosas con competencia y con una comprensión escasísima (pensemos en los castillos de las termitas o en los antílopes saltarines), que se nos plantea un nuevo misterio: ¿para qué sirve la comprensión? y ¿cómo puede surgir una mente como la de Bach o la de Gaudí? Al echar un vistazo más detallado al diseño de los ordenadores y cómo están diseñados para usar información para realizar tareas antes reservadas a pensadores humanos con comprensión, pudimos aclarar la diferencia entre procesos de diseño «ascendentes» o «de abajo arriba», representados por las termitas y por la selección natural, y los procesos de diseño inteligente «descendentes» o «de arriba abajo». Esto nos llevó a la idea de la información como diseño que vale la pena robar… o copiar, o comprar. La excelente teoría de la información de Shannon clarifica la idea básica (una diferencia que marca la diferencia) y le aporta un fundamento teórico sólido y maneras de medir la información. Pero necesitamos ir más allá si queremos ver por qué estas diferencias son tan valiosas, por qué vale tanto la pena medirlas, en primer lugar.

Los diversos procesos de evolución darwiniana no son todos iguales, algunos son más darwinianos que otros procesos que son igual de reales e igual de relevantes en sus propios nichos, de modo que es importante ser darwiniano sobre el darwinismo. Los espacios darwinianos de Godfrey-Smith son un buen recurso de pensamiento para ayudarnos a visualizar las relaciones entre las formas en que evolucionan diferentes especies y también la forma en que evoluciona la propia evolución, con algunas estirpes que muestran una desdarwinización a medida que pasa el tiempo.

Volviendo al misterio de cómo los cerebros formados por miles de millones de neuronas sin ningún sistema de control de arriba abajo pueden llegar a formar mentes al estilo de los humanos, exploramos la hipótesis de un control descentralizado y distribuido por parte de neuronas pertrechadas para valerse por sí mismas, e incluimos la posibilidad de neuronas salvajes, liberadas de su papel anterior como esclavos dóciles y domesticados bajo la presión selectiva creada por un nuevo rasgo ambiental: los invasores culturales. Las palabras que luchan por reproducirse, así como otros memes, provocarían adaptaciones tales como reestructuraciones del cerebro en una respuesta coevolutiva. Una vez quedó garantizada la transmisión cultural como principal innovación

cultural de nuestra especie, esta transmisión no solo provocó cambios importantes en la arquitectura neuronal, sino que también añadió novedades en el entorno (en forma de ofrecimientos gibsonianos) que enriquecieron las ontologías de los seres humanos y, a su vez, generaron más presiones selectivas que favorecieron las adaptaciones (recursos de pensamiento) que permitían controlar y rastrear todas estas nuevas oportunidades. La evolución cultural también evolucionó, alejándose de búsquedas aleatorias y sin dirección y acercándose a procesos de diseño más eficaces, con previsión, con propósitos y con una dependencia de la comprensión de los agentes: los diseñadores inteligentes. Para la comprensión humana se necesita un amplísimo abanico de recursos de pensamiento. La evolución cultural se desdarwinizó con los productos que ella misma había creado.

Esta perspectiva nos permite ver la imagen manifiesta, siguiendo la útil terminología de Wilfrid Sellars, como un tipo especial de artefacto, en parte diseñado genéticamente y en parte diseñado culturalmente, una ilusión de usuario muy eficaz para ayudar a los apremiados organismos a apañarse con destreza en la vida, valiéndose de (hiper)simplificaciones que crean una imagen del mundo en que vivimos que está en cierto conflicto con la imagen científica a la cual debemos recurrir para explicar el surgimiento de la imagen manifiesta. He aquí otra revolucionaria inversión del razonamiento, que aparece en la descripción de Hume de nuestro conocimiento de la causalidad. Así, podemos ver la conciencia humana como una ilusión de usuario, no representada en el teatro cartesiano (que no existe) sino formada por las actividades representacionales del cerebro junto con las reacciones adecuadas a tales actividades («y ahora, ¿qué?»).

Con esto se puede soldar la fractura, suturar la herida cartesiana, pero en este momento solo queda claro un esbozo de esta importante unificación. Aun así, este borrador tiene el detalle suficiente para mostrar que las mentes humanas, por muy inteligentes y comprendedoras que sean, no son los sistemas cognitivos más potentes imaginables; nuestros diseñadores inteligentes han hecho progresos increíbles creando sistemas de aprendizaje automático que usan procesos de abajo arriba para demostrar una vez más la Segunda Regla de Orgel: la evolución es más lista que tú. En cuanto apreciamos la universalidad de la perspectiva darwiniana, nos damos cuenta de que nuestro estado actual, como individuos y como sociedad, es imperfecto y temporal. Quizá algún día

devolvamos el planeta a nuestros parientes bacterianos y su estilo modesto, de abajo arriba, de mejoras de diseño; o podemos continuar progresando, en un entorno que hemos creado con ayuda de artefactos que hacen la mayoría del trabajo pesado cognitivo a su manera, en una era de posdiseño inteligente. No hay solo coevolución entre memes y genes; también hay codependencia entre las capacidades de razonamiento de arriba abajo de nuestras mentes y los talentos de abajo arriba, sin comprensión, de nuestros cerebros animales. Si nuestro futuro sigue la senda de nuestro pasado (una cosa que, en parte, depende de nosotros), nuestras inteligencias artificiales continuarán dependiendo de nosotros a medida que nosotros, cautelosamente, nos volvemos más dependientes de ellas.

ANEXO

ALGUNOS ANTECEDENTES

Las ideas y los argumentos de este libro son la parte más visible de un conjunto de argumentos que he ido elaborando durante medio siglo. He intentado escribir el libro de modo que el lector no necesite ningún conocimiento de mis obras previas y a las cuales solo he hecho unas pocas referencias. Los lectores, especialmente si aún no están convencidos, pueden preguntarse si las afirmaciones que expongo tienen algún respaldo adicional, y la respuesta es que sí. Algunos también querrán saber la historia que hay detrás de los argumentos planteados, para poder reseguir la historia de las diversas afirmaciones, que han ido cambiando. Este anexo proporciona referencias a todas las obras en las que me baso y también a algunas de las obras críticas más útiles que han provocado.

Pero antes de pasar a estos antecedentes, dejadme cumplir la promesa de la nota 6 del capítulo 13 (p. 390) sobre Peter Carruthers y mostrar algunas obras muy relevantes que he descubierto pero que aún no he tenido tiempo de asimilar adecuadamente para incluirlas en este libro. Además de The Centered Mind de Carruthers y sus obras anteriores, me gustaría recomendar Murray Shanahan, Embodiment and the Inner Life, 2010; Radu Bogdan, Our Own Minds: Sociocultural Grounds for Self-Consciousness, 2010, y Mindvaults: Sociocultural Grounds for Pretending and Imagining, 2013; Andy Clark, Surfing Uncertainty, 2015; y Olivier Morin, How Traditions Live and Die, 2016. En un mundo ideal, hubiera tenido tiempo de incluir todas estas obras en mis trabajos previos a la redacción del libro y, sin duda, el libro se habría beneficiado de ello, pero hay que acabar en algún momento y concentrarse en escribir. Espero poder comentar todas estas obras (y más) en un futuro no muy lejano, probablemente en mi sitio web.

CAPÍTULO 1

«los términos romántico y aguafiestas».

Analicé por primera vez el pulso entre románticos y aguafiestas, con respecto a las mentes animales, en «Intentional Systems in Cognitive Ethology: The Panglossian Paradigm Defended», un artículo destacado en Behavioral and Brain Sciences (1983). Este ensayo fue también el inicio de mi crítica del ensayo sobre el «paradigma panglossiano» de Gould y Lewontin, que elaboré durante más de doce años en un ataque cada vez más detallado contra la incorrecta dirección tomada por la obra de Gould. Véase, por ejemplo, mi capítulo sobre Gould en La peligrosa idea de Darwin (1995) y los intercambios subsiguientes con Gould en la New York Review of Books (1997, 1997b) y en Dennett, «Darwinian Fundamentalism» (1997).

«un precipicio por el borde del cual puedes empujar a tus oponentes».

Describí este precipicio en «Current Issues in the Philosophy of Mind» (1978b).

«intentar responder preguntas retóricas».

Sobre las preguntas retóricas como reveladores puntos débiles en los argumentos, véase Bombas de intuición (2013).

«exagerándolo hasta convertirlo en algo profundo y metafísico».

Humphrey y Dennett, «Speaking for Our Selves» 1989 está reimpreso en Dennett, Brainchildren (1998).

CAPÍTULO 2

«un diseñador inteligente de otra galaxia».

Véase Dennett «Passing the Buck to Biology» en Behavioral and Brain Sciences (1980).

«El punto de vista de la ingeniería inversa es ubicuo en biología».

La biología como una especie de ingeniería inversa se trata en profundidad en el capítulo 8 de La peligrosa idea de Darwin (1995).

CAPÍTULO 3

«el enfoque de diseño».

El enfoque de diseño o actitud de diseño fue presentado en «Intentional Systems» en 1971, y desde entonces ha sido un protagonista destacado de mis obras. Véase, especialmente, Brainstorms (1978), «Three Kinds of Intentional Psychology» (1981a), «Making Sense of Ourselves» (1981b) y The Intentional Stance (1987).

«vestigial».

Podéis ver una discusión de algunos ejemplos de estructuras de artefactos que han perdido su razón de existir en La peligrosa idea de Darwin (1995).

«ciclos».

La afirmación de que los ciclos (no solo los reproductivos) pueden aumentar la probabilidad de que suceda algo nuevo era el tema de mi ensayo para la pregunta de 2011 de Edge.org: «¿Qué concepto científico podría mejorar el juego de herramientas cognitivo de todo el mundo?», posteriormente publicado (2012c) en John Brockman, ed., This Will Make You Smarter.

«Primer Mamífero».

Analizo la afirmación de David Sanford acerca de la inexistencia de mamíferos en La peligrosa idea de Darwin (1995) y luego en Bombas de intuición (2013).

«motivaciones flotantes».

Introduje el término motivaciones flotantes en mi «Intentional Systems in Cognitive Ethology» (1983) y luego he profundizado el concepto en La peligrosa idea de Darwin (1995), en Bombas de intuición (2013) y en otras obras.

CAPÍTULO 4

Este capítulo se basa, con numerosas revisiones, en mi artículo en los PNAS (2009), «Darwin's 'Strange Inversion of Reasoning'» y en «Turing's 'Strange Inversion of Reasoning'» (2013d).

CAPÍTULO 5

«El físico y matemático Roger Penrose (1985) es el representante más ilustre de esta postura».

He analizado en profundidad el argumento de Penrose contra la IA (y, de hecho, ¡contra la selección natural!) en «Murmurs in the Cathedral», mi reseña del libro de Penrose en el Times Literary Supplement (1989), en el capítulo 15 de La peligrosa idea de Darwin (1995) y en Brainchildren (1998), especialmente en los capítulos de la parte II.

«optimización».

Para más información sobre las suposiciones de optimización en biología evolutiva, véase «Intentional Systems in Cognitive Ethology», 1983, y La peligrosa idea de Darwin (1995), así como mis trabajos sobre el enfoque de diseño, el enfoque intencional y el enfoque físico, citados antes para el capítulo 3.

CAPÍTULO 6

«He aquí un ejemplo algo rebuscado».

El ejemplo de Jacques y Sherlock es una adaptación de «Three Kinds of Intentional Psychology» (1981) y también aparece en Bombas de intuición (2013). En las versiones previas, comentaba explícitamente las creencias de un sistema intencional y utilizaba el ejemplo para evitar resbalar hacia la tentadora conclusión de que tenía que haber un «lenguaje de pensamiento» en el que, de algún modo, estarían codificadas todas las creencias, un tema que ha dado muchas vueltas en mi obra. Véase también La actitud intencional (1987) y Westbury y Dennett «Mining the Past to Construct the Future: Memory and Belief as Forms of Knowledge» (1998). Hoy en día generalizo esta idea hablando de información. La información acerca de la viscosidad del medio en que vuelan las aves encarnada en el diseño de las alas de las aves (y de sus controles) no tiene que estar codificada en este diseño. Al revisar y mejorar el concepto de enfoque intencional o actitud intencional espero haber aclarado y ampliado esta idea: la utilidad de adoptar una perspectiva informacional/intencional no depende de la existencia de un sistema de representación para codificar este valioso producto, la información semántica.

«una excesiva transparencia».

Sobre los peligros de la transparencia, véase Bombas de intuición (2013) y Dennett y Roy (2015).

«¿Cui bono? ¿A quién beneficia?».

Discutí por primera vez la importancia de hacerse esta pregunta en La peligrosa idea de Darwin (1995).

«los Buenos Trucos».

Defino y comento los Buenos Trucos en La peligrosa idea de Darwin (1995), así como los «cuentos de así fue».

«se establece una confianza».

Para un punto de vista diferente acerca de la confianza interna, véase McKay y Dennett «The Evolution of Misbelief», Behavioral and Brain Sciences (2009).

«patrón real de información potencialmente útil».

Véase «Real Patterns» (1991b) para un extenso análisis de la relación entre ontología y patrones discernibles.

«un carro con ruedas de radios».

El ejemplo de este carro está comentado en La conciencia explicada (1991) y en La peligrosa idea de Darwin (1995).

CAPÍTULO 7

«Siempre que se dan estos tres factores».

Este es el «algoritmo» darwiniano que comento con detalle en Dennett (1995).

«huteritas».

En el capítulo «Algunas variaciones del reduccionismo ético insaciable» de La peligrosa idea de Darwin (1995, pp. 778-803) presento una discusión crítica del análisis de Sober y Wilson de los huteritas.

CAPÍTULO 8

«Los cerebros procesan en paralelo».

Paul Churchland y yo tenemos un debate inacabable sobre la potencia de las arquitecturas en paralelo. Como muestra el título de mi último capítulo, las discrepancias están disminuyendo: «Two Steps Closer on Consciousness» (2006d). Véase también mi discusión en La evolución de la libertad (2003).

«en su arduo e importante libro».

Véase mi reseña del libro de Deacon: «Aching Voids and Making Voids» (2013b).

«Tecumseh Fitch».

Los seis párrafos previos están extraídos, con algunas revisiones, de mi contribución a la pregunta de Edge.org de 2008: ¿En qué ha cambiado tu mente, y por qué?

«¿Cómo recoge el cerebro los ofrecimientos?».

Este apartado se ha beneficiado de las discusiones con David Baßler y, con algunas revisiones, se basa en «Why and How Does Consciousness Seem the Way It Seems?» y «How Our Belief in Qualia Evolved, and Why We Care So Much—A Reply to David H. Baßler», Open MIND, Metzinger 2015 y en «Expecting Ourselves to Expect: The Bayesian Brain as a Projector», mi comentario en Behavioral and Brain Sciences (2013c) sobre Clark (2013).

«modelos de competencia frente a modelos de ejecución».

En las ciencias cognitivas, han sido influyentes distinciones relacionadas con esta (pero mejores): David Marr y Tommy Poggio (1976) propusieron tres niveles: computacional, algorítmico e implementativo; Allen Newell (1982) propuso el nivel de conocimiento, el nivel simbólico y el nivel de hardware; mis tres «enfoques» o «actitudes» (intencional, de diseño y físico) fueron propuestos un poco antes (1971) y eran algo más generales. Cada versión de lo que es, fundamentalmente, la misma idea tiene sus inconvenientes: el término computacional de Marr y Poggio es equívoco para el nivel de las especificaciones; el término simbólico de Newell expresa un sesgo de la GOFAI y mi enfoque intencional aporta a la ciencia cognitiva una pizca de jerga filosófica controvertida y a menudo mal comprendida.

«postular algún "tejido milagroso"».

Este tejido milagroso se menciona por primera vez en una obra impresa en «Cognitive Wheels» (1984b).

CAPÍTULO 9

«Avital y Jablonka».

Para obtener más información sobre el libro de Avital y Jablonka, véase mi reseña en el Journal of Evolutionary Biology (2002b).

«¡Muera la democracia!».

Las vicisitudes para «descifrar» las actividades del cerebro como forma de atribuir creencias ha sido uno de mis temas preferidos desde 1975. Véase «Brain Writing and Mind Reading», reimpreso en Brainstorms (1978).

CAPÍTULO 10

«versiones novedosas o contextos más amplios».

Mi interés en la etología cognitiva y los problemas de atribuir estados intencionales de alto nivel a animales no humanos se remonta a mi artículo destacado en Behavioral and Brain Sciences de 1983, «Intentional Systems in Cognitive Ethology: The 'Panglossian Paradigm' Defended» y ha sido ampliado desde entonces. Véase, por ejemplo, «Cognitive Ethology: Hunting for Bargains or a Wild Goose Chase?» (1989), «Do Animals Have Beliefs?» (1995b) y «Animal Consciousness: What Matters and Why» (1995b).

CAPÍTULO 11

«Los patrones».

En Dennett (1991b) analizo patrones reales.

«thinko».

El concepto de thinko lo presenté en «From Typo to Thinko: When Evolution Graduated to Semantic Norms» (2006c).

CAPÍTULO 12

«véase Humphrey 1998».

Contesto a Nicholas Humphrey, «Cave Art, Autism, and the Evolution of the Human Mind» en el Cambridge Archaeological Journal (1998b).

«Quizá solo seamos monos con cerebros manipulados por memes».

Véanse los capítulos sobre memes en La conciencia explicada (1991), La peligrosa idea de Darwin (1995) y «The evolution of culture» (2001b).

«tal como la hormiga que sube por una brizna de hierba es la forma que tienen los trematodos».

Sobre este tema, véase «The evolution of culture» (2001b) y Romper el hechizo: la religión como un fenómeno natural (2006).

«las llamadas de aviso de los cercopitecos verdes».

Mi artículo sobre etología cognitiva en Behavioral and Brain Sciences (1983) comentaba las dificultades para diferenciar anécdotas de pruebas científicas de inteligencia animal. «Beliefs about beliefs» (1978d), mi comentario sobre el artículo destacado de Premack y Woodruff en Behavioral and Brain Sciences, «Does the Chimpanzee Have a Theory of Mind?» (1978), fue seguido por una gran diversidad de estudios, de campo y de laboratorio, acerca de signos de intencionalidad de alto nivel (como las pruebas «de falsa creencia») en animales y niños. Véase, por ejemplo, Whiten y Byrne (1988) Machiavellian Intelligence y Machiavellian Intelligence II: Extensions and Evaluations (1997), así como trabajos más recientes de Tomasello, Call, Povinelli y muchos otros. En Tipos de mentes (1996) comentaba la diferencia entre tener creencias acerca de creencias y pensar sobre pensar, y también repasaba parte de las publicaciones sobre el tema. Véanse también mis comentarios en «The Selfish Gene as a Philosophical Essay» (2006b).

«cualquier intento de explicación del DAL mediante la selección natural».

En «Passing the Buck to Biology» (Behavioral and Brain Sciences, 1980) ofrecí lo que pensaba que era una amistosa corrección al artículo de Chomsky, y me quedé perplejo con su intransigente respuesta. Yo creía que la idea básica del DAL era afirmar que era la I+D darwiniana la

responsable de diseñar una gramática, más que la skinneriana, popperiana o gregoriana, pero estaba equivocado.

CAPÍTULO 13

«Un orden que está presente».

Comento este fragmento de Anscombe en «Real Patterns» (1991b, p. 43).

«un andamio conversacional que ayuda a los niños a acostumbrarse a su lengua».

Véase Griffin y Dennett (2008) acerca del andamiaje en el contexto del autismo.

«yo la puedo metahacer».

Cito la ocurrencia de Hofstadter «cualquier cosa que puedas hacer, yo la puedo metahacer» en «Hofstadter's Quest» (1995b). Hofstadter afirma que soy yo el autor de la ocurrencia, pero es irrelevante a quién se le haya ocurrido primero. En 2004, a la pregunta de Edge.org «¿Cuál es tu ley?» Charles Simonyi promulgó la Ley de Simonyi: cualquier cosa que se pueda hacer, se puede metahacer. Las buenas ideas son Buenos Trucos, que suelen reinventarse con frecuencia.

«máquina virtual Java».

Para una introducción más detallada a las máquinas virtuales y a su relación con el hardware y el software, véase Bombas de intuición (2013, capítulo 4 «Intermedio sobre computadoras»), que ofrece una suerte de manual básico.

«Consideramos a Oscar Wilde».

Comenté por primera vez el ejemplo de Wilde generando trabajosamente sus ocurrencias en «Get Real» (1994).

«Problema Difícil».

En Dulces sueños: obstáculos filosóficos para una ciencia de la conciencia (2005) y en Bombas de intuición (2013) comento extensamente el Problema Difícil y la «magia de verdad» de la conciencia.

«Spakesheare».

El experimento mental de Spakesheare y Spamlet apareció por primera vez en mi discurso presidencial ante la Eastern Division de la American Philosophical Association, publicado como «In Darwin's Wake, Where Am I?» (Proc. APA 2000).

«mente extendida».

Comento la importancia de usar marcas en el entorno en Tipos de mentes (1996) y en «Learning and Labeling» (1993).

CAPÍTULO 14

«científico "marciano"».

La idea de que la conciencia sería evidente para los científicos marcianos aparece por primera vez en Dulces sueños (2003).

«¿ser como qué?».

Véanse las divertidas y perspicaces reflexiones por parte de Hofstadter en Hofstadter y Dennet, The Mind's I (1981), pp. 403-414.

«no tardará en extinguirse».

Para una discusión acerca de la teoría de juegos y la privacidad, véase La evolución de la libertad (2003).

«comunicación específicamente lingüística».

Los dos párrafos finales del apartado se basan, con muchos cambios, en «My Body Has a Mind of Its Own» (2007c) y «The Evolution of 'Why'» (2010), que presentan más comentarios de los orígenes de esta postura en las obras de McFarland (1989, 1989b), Drescher (1991), Brandom (1994), Haugeland (1998) y Millikan (2000b). Véase también mi reseña (2009b) del libro de McFarland de 2009 Guilty Robots, Happy Dogs.

«más "mísmico"».

En inglés, selfy, cuya acuñación a partir de sexy se comenta en «La conciencia explicada» (1991).

«Tal como lo expresa Daniel Wegner».

En La evolución de la libertad presento una extensa discusión del punto de vista de Wegner del libre albedrío como una ilusión.

«heterofenomenología».

Estos párrafos se basan en buena parte en «Heterophenomenology Reconsidered» (2007b), donde la discusión es mucho más profunda; véase, en particular, la respuesta a Schwitzgebel (2007) y, en la nota 6, a Siewert (2007), cuyo pensamiento me ayudó a reformular estas ideas. La heterofenomenología (1982, 1991, 2003c, 2005, 2007c) nos permite rastrear los ofrecimientos, las «cosas» en el «mundo nocional» de una persona sin refrendar esa ontología como parte de nuestra ontología científica.

n. 2: «en una imagen que se puede consultar».

Sobre la nota 2 acerca de investigación en imaginería mental: Shepard y Metzler (1971) fue la brillante chispa que provocó una explosión de trabajos, con Pylyshyn (1973) y Kosslyn (1975) tendiendo la mesa para las controversias posteriores (véase, p. ej., Kosslyn 1980, Kosslyn et al. 1979, Kosslyn et al. 2001, Pylyshyn 2002). Para mis reflexiones sobre el tema, véase «Contenido y conciencia» (1969), «Brainstorms» (1978, cap. 9 y 10), «La conciencia explicada» y «Does Your Brain Use the Images in It, and If So, How?» (comentario en Pylyshyn 2002) en Behavioral and Brain Sciences 2002.

«la extraña inversión del razonamiento de Hume».

Los párrafos sobre esta inversión se basan, con revisiones, en mi comentario sobre Clark, Behavioral and Brain Sciences, 2013. En Hurley, Dennett y Adams, Inside Jokes, se puede hallar una presentación anterior de algunas de estas ideas.

«la Pregunta Difícil».

Para una extensa discusión de la Pregunta Difícil, véase «La conciencia explicada» (1991); para un ejemplo de hacerse la Pregunta Difícil y empezar a responderla, véase «The Friar's Fringe of Consciousness» (2015b).

«no se pueden satisfacer ambas posturas».

Puedo ofrecer bombas de intuición para que mi afirmación, como mínimo, pueda ser contemplada por los que, a primera vista, la encuentran incomprensible. Véase, en particular, «The Tuned Deck» en «Explaining the 'Magic' of Consciousness» (2002c) (de donde procede parte del material de los párrafos anteriores), Dulces sueños (2005) y Bombas de intuición (2013).

«identificación por el contenido».

Para obtener más información sobre Bennett y esta identificación por el contenido, véase «Jonathan Bennett's Rationality» (próxima publicación). Véase también Azzouni (2013) acerca de cómo moran los significados en nuestra fenomenología aun no existiendo.

«libre albedrío».

Sam Harris (2012) y otros (como Coyne 2012) han sostenido recientemente que la ciencia demuestra que no existe el libre albedrío y que esto tiene consecuencias revolucionarias (aunque no especificadas) para nuestras políticas de moralidad y la ley, el crimen, el castigo y las recompensas. Ignoran el hecho de que la imagen manifiesta es donde vivimos y es lo que nos resulta relevante. Desde la perspectiva de la imagen científica, el color es una ilusión, pero no por eso nos resulta irrelevante y por ello gastamos tiempo y esfuerzo en rodearnos de los colores que apreciamos. Es como si Harris y los demás sostuvieran que el dinero es una ilusión y, en consecuencia, que debemos abandonarlo. No creo que dejen sin más todo su ilusorio dinero, ni creo que eludan la responsabilidad de sus acciones intencionales, de modo que no está claro qué es lo que creen que están demostrando. Para obtener más información sobre el libre albedrío, véanse mis libros Elbow Room: The Varieties of Free Will Worth Wanting y La evolución de la libertad (2003), así como mi conferencia Erasmus «Erasmus: Sometimes a Spin Doctor Is Right» (2012b); sobre Harris: «Reflections on Free Will»

(2014b) y «Daniel Dennett on Free Will Worth Wanting» en el podcast Philosophy Bites Again (2014).

CAPÍTULO 15

«Emerson P. Pugh».

Citado por George E. Pugh en The Biological Origin of Human Values (1978, p. 154). En una nota, el autor afirma que esta cita es de su padre, Emerson Pugh, hacia 1938. La cita también se suele atribuir a Lyall Watson, nacido en 1939. http://todayinsci.com/P/Pugh_Emerson/PughEmerson-Quotations.htm.

«un libro con muchos autores y muchos volúmenes».

En el capítulo 8 de Romper el hechizo (2006) comento la división del trabajo entre científicos de diferentes campos que se basan en los conocimientos de los otros («¡ellos comprenden y nosotros creemos!») y señalo que en la teología se suele destacar que nadie entiende los términos de la discusión.

«avanzados nanorrobots evolucionados».

Los robots evolucionados a escala macroscópica también han logrado resultados espectaculares en campos muy simplificados y también he comentado a menudo el trabajo en robótica evolutiva de Inman Harvey y Phil Husbands en Sussex (p. ej., Harvey et al. 1997), pero no lo he discutido en ninguna obra publicada.

«David Cope».

La obra Virtual Music de Cope (2001) incluye ensayos de Douglas Hofstadter, compositores, musicólogos y míos: «Collision Detection,

Muselot, and Scribble: Some Reflections on Creativity». Los ensayos están repletos de llamativas observaciones y ejemplos y la obra incluye partituras y un CD.

«no depende del sustrato».

Véase La peligrosa idea de Darwin (1995) acerca de la independencia del sustrato.

«analogizadores».

Véanse también las numerosas obras de Douglas Hofstadter acerca de la importancia de la búsqueda de analogías, especialmente Metamagical Themas (1985), Fluid Concepts and Creative Analogies (1995) y L'Analogie. Coeur de la pensée, escrito junto con Emmanuel Sander.

«Los arquitectos».

Para ver algunos ejemplos del uso de algoritmos genéticos en arquitectura, véase Sullivan-Fedock (2011), Asadi et al. (2014) y Yu et al. (2014) y, para la optimización de los distritos electorales, Chou et al. (2012).

«la capacidad de autoobservarse».

Comento esta estrategia de diseño en un artículo no publicado, «A Route to Intelligence: Oversimplify and Self-monitor» (1984c), que puede hallarse en mi sitio web: http://ase.tufts.edu/cogstud/dennett/recent.html.

«el clásico test de Turing».

Para un análisis y defensa del test de Turing como prueba genuina de comprensión, véase mi artículo «Can Machines Think?» (1985), reimpreso en Brainchildren, con dos posfacios (1985) y (1997); «El pensamiento veloz» en La actitud intencional (1987) y, en particular, «La habitación china» en Bombas de intuición (2013), donde discuto ejemplos de las capas cognitivas que deben intercambiarse en una conversación.

«Debner y Jacoby».

Para consultar más información sobre estos experimentos, véase mi «Are We Explaining Consciousness Yet?» (2001c) y también Dehaene y Naccache (2001), Smith y Merikle (1999), comentado en Merikle et al. (2001).

«una teoría lo bastante detallada y dinámica de agentes con imaginación».

Comento las perspectivas de una teoría o modelo tan potente de un agente inteligente, y señalo una ambigüedad clave en el test de Turing original, en una entrevista con Jimmy So acerca de las implicaciones de Her, en «Can Robots Fall in Love» (2013), The Daily Beast, http://www.thedailybeast.com/articles/2013/12/31/can-robots-fall-in-love-and-why-would-they.html.

«una posibilidad práctica insignificante».

Cuando explico por qué pensaba que la IA era posible en principio pero imposible en la práctica, a menudo la comparaba con la tarea de construir un pájaro robótico que no pesara más que un mirlo, que pudiera cazar insectos volando y que pudiera posarse en una ramita. No había ningún misterio cósmico en tal pájaro, afirmaba, pero la ingeniería necesaria para hacerlo realidad costaría más de doce proyectos Manhattan, y ¿para qué? Podemos aprender todo lo que necesitamos saber sobre los principios del vuelo, incluso del vuelo de las aves,

elaborando modelos más simples con los cuales comprobar nuestras teorías, y con una fracción mínima del coste. Mucha gente me ha interpelado últimamente con las últimas novedades de drones en miniatura, más pequeños que mirlos, tan diminutos como insectos, y me piden si quiero reconsiderar mi analogía. No, porque no son robots autónomos, sino marionetas electrónicas, y tampoco pueden cazar moscas volando ni posarse sobre ramitas. Quizá algún día se podrá hacer algo así, pero solo si la DARPA invierte miles de millones de dólares. He comentado la Singularidad en « The Mystery of David Chalmers » (2012) y en «The Singularity—An Urban Legend?» (2015).

«cediéndoles prematuramente una autoridad».

Amplío esta preocupación en «Information, Technology, and the Virtues of Ignorance» (Daedalus, 1986), reimpreso en Brainchildren (1998) y en «The Singularity—An Urban Legend?» (2015).

«cuando el coche se estropea».

Estos párrafos se basan, con revisiones, en mi prefacio de la segunda edición de Seabright, The Company of Strangers, 2010.

NOTAS

CAPÍTULO 1

1. Uno de mis favoritos es el siguiente: cuatro personas llegan a un río, de noche. Hay un pequeño puente que solo puede soportar a dos personas a la vez. Tienen una linterna y, como es de noche, se tiene que usar la linterna para pasar el puente. La persona A puede pasar el puente en un minuto; la persona B, en dos minutos; la persona C, en cinco minutos y la persona D, en ocho; y cuando dos personas pasan a la vez por el puente, deben moverse a la velocidad de la más lenta. La pregunta es: ¿pueden pasar todos ellos el puente en quince minutos o menos?

2. Lane (2015) presenta un repaso y una actualización fascinantes de la historia del origen endosimbiótico de los eucariotas que he transmitido durante los últimos veinte años. Actualmente parece bastante claro que Adán y Eva fueron una bacteria y una arquea, no dos bacterias diferentes, como había afirmado a menudo.

3. Como este libro es la culminación de medio siglo de trabajo en estos temas, las prácticas académicas habituales de citación me obligarían a rociar estas páginas con docenas de interrupciones del tipo «(Dennett 1971, 1991, 2013)», pero esta abundante autocitación resultaría equívoca. Mis ideas han sido esculpidas por cientos de pensadores y he intentado citar las fuentes básicas de todas las ideas que se comentan a medida que surgen, mientras que la información sobre dónde he ido ampliando información sobre estos temas la he recluido en el anexo «Algunos antecedentes» (p. 367) para comodidad de todos aquellos que quieran saber cómo han ido progresando los diversos razonamientos.

4. En su libro de 1979, Scientific Realism and the Plasticity of Mind, Churchland incluye instrucciones y un diagrama que ayudarán a disfrutar de este efecto.

5. Hace algunos años, trabajando con Nick Humphrey en lo que entonces se llamaba trastorno de personalidad múltiple (hoy, trastorno disociativo), descubrí la tentación casi irresistible (también en Nick y en mí mismo) de exagerar cualquier cosa que te sorprende como importante e inesperada. Supongo que cuando los seres humanos nos encontramos con algo extraño e inquietante, nuestros intentos por describirnos lo que estamos experimentando tienen tendencia a pecar de exageración, quizá por un deseo subliminal de impresionarnos con el imperativo de que si lo ignoramos será a nuestra cuenta y riesgo y de que debemos llegar al fondo del tema.

CAPÍTULO 2

1. No se trata de una cita de Hermann Göring (ni tampoco de Heinrich Himmler). Según la Wikipedia, esta frase, a menudo atribuida incorrectamente, procede de un diálogo de una obra de teatro pronazi de Hanns Johst: «Wenn ich Kultur höre [...] entsichere ich meinen Browning!»

2. En este libro diré muchas cosas acerca del diseño inteligente (realizado por seres humanos), pero prácticamente nada más acerca del Diseño Inteligente, la última oleada de la propaganda creacionista. No merece más refutaciones.

3. Nikolai Renedo me ha sugerido que el mensaje neto del ensayo de Gould y Lewontin es «ten cuidado con los gambitos», lo que sin duda es un buen consejo para cualquier adaptacionista. Si es esto lo que realmente querían transmitir Gould y Lewontin, no consiguieron expresarlo claramente a su audiencia, tanto especializada como profana, ya que se mantiene la opinión de que era una atenuación bien documentada del adaptacionismo como rasgo central del pensamiento evolutivo.

CAPÍTULO 3

1. Biólogos y filósofos han escrito a menudo acerca de la idea de función, y aunque hay desavenencias persistentes sobre cómo se debe utilizar, hay un cierto consenso en que las consideraciones evolutivas sirven, de una manera u otra, para las funciones naturales, mientras que los hechos acerca de las capacidades pasadas y actuales fijan los atributos de las funciones a los artefactos. Para una buena antología de los mejores trabajos de biólogos y filósofos, véase Allen, Bekoff y Lauder (1998).

2. Por ejemplo, el biólogo Shirley Tilghman, en la Conferencia Watson de 2003, dijo: «sin duda, lo que es evidente a simple vista cuando observamos cualquier parte de estos dos genomas que hemos comparado, es que la evolución se ha esforzado, conservando mucho más del genoma de lo que podemos explicar a partir de los genes y de sus elementos reguladores asociados [...] Los científicos deberían aprovecharlo al máximo para intentar comprender lo que la evolución tenía pensado al fijarse tanto en estos pequeños fragmentos de ADN».

3. Los filósofos más escépticos acerca de mi intransigencia en este tema podrían leer el reciente libro de T. M. Scanlon, Being Realistic about Reasons (2014), donde hallarán un exhaustivo y agotador estudio de los problemas a los que uno se enfrenta si ignora las razones de ingeniería y se concentra en disponer de razones morales para las acciones.

CAPÍTULO 4

1. Durante más de treinta años, he llamado a este autor, incorrectamente, Robert Beverley MacKenzie; agradezco a los revisores de Norton haberme sacado de mi error.

2. En la jerga habitual, esto se denomina tesis de Church-Turing, formulada por el lógico Alonzo Church: «todos los procedimientos eficaces son computables Turing», si bien muchos de ellos no son factibles porque tardan demasiado. Como nuestra comprensión de qué es

un procedimiento eficaz (básicamente, un algoritmo o programa informático) es intuitiva, esta tesis no se puede demostrar, pero se acepta casi universalmente; tanto es así que la computabilidad Turing se suele usar como una definición operativa aceptable de eficacia.

3. Gaudí murió en 1926, pero dejó dibujos, instrucciones y modelos que siguen guiando la construcción de su obra inacabada; Turing abandonó el NPL antes de que se completara el Pilot ACE, pero también dejó representaciones del artefacto para orientar en su fabricación.

4. En realidad, Sellars diferenciaba una «concepción del hombre en el mundo precientífica, acrítica e ingenua [que] podría denominarse la imagen original» (1962) de la imagen manifiesta, un «refinamiento o sofisticación» de esa imagen original. A lo que quería llegar con esta distinción es que, durante milenios, los filósofos han estado reflexionando críticamente acerca de la concepción ingenua, de modo que la imagen manifiesta no era solo metafísica popular.

5. Reconozco que esta afirmación es algo rotunda; no veo razón alguna para creer que los árboles o las bacterias tengan sistemas de control más parecidos a nuestras mentes que los sistemas de control de un ascensor, pero admito que es posible que así sea. Considero que la posibilidad es insignificante, un riesgo no nulo que estoy dispuesto a asumir.

6. El proyecto Cyc de Douglas Lenat es un intento, actualmente en curso, de crear una inteligencia artificial de este tipo; después de treinta años de trabajo con un código escrito por centenares de programadores (conocidos como Cyclists), su memoria dispone de más de un millón de conceptos definidos a mano.

CAPÍTULO 5

1. Hay una considerable controversia sobre el uso del término optimizar al referirse a productos «lo bastante buenos» de la selección natural. El proceso de selección natural no puede tener en cuenta todas las cosas posibles y siempre se halla en medio de un proceso de rediseño, de modo

que nada garantiza que halle la solución óptima de cualquier problema específico de diseño que se plantee; aun así, lo hace asombrosamente bien, en general mejor que los diseñadores inteligentes humanos que se esfuerzan por lograr diseños óptimos.

2. El mítico creador de software Charles Simonyi, autor principal del Microsoft Word, ha dedicado más de veinte años a la tarea de crear lo que él denomina software intencional, que, en el mejor de los casos, resolvería este problema o un importante subconjunto del mismo. El hecho de que varias décadas de trabajo de gran calidad por parte de un equipo de ingenieros de software no haya dado resultados tangibles es una buena muestra de la dificultad del problema.

3. La evolución explora los «posibles adyacentes» (véase Kauffman 2003).

4. ¿Cómo puedo decir que la evolución, uno de cuyos rasgos más famosos es la falta de previsión, puede (o no) predecir algo? Podemos aceptar este uso práctico del enfoque intencional aplicado a la evolución diciendo, quizá de una forma menos memorable e instructiva, que los entornos muy variables no disponen de información acerca de los entornos futuros que la selección natural pueda aprovechar (ciegamente). Véase también el capítulo 6.

5. El término historia whig se refiere a la interpretación de la historia como una secuencia de progreso, que suele justificar la cadena de eventos que llevan al privilegiado punto de vista del intérprete. Para aplicaciones del término al adaptacionismo en biología evolutiva, tanto a favor como en contra, véase Cronin (1992) y Griffiths (1995).

CAPÍTULO 6

1. En Colgate y Ziock (2010) se presenta un resumen conciso y útil de la historia de las definiciones de información a partir de los trabajos de Shannon y Weaver.

2. Giulio Tononi (2008) ha propuesto una teoría matemática de la conciencia como «información integrada» que usa la teoría de la información de Shannon de un modo novedoso y tiene un papel muy limitado para el «acerca de»: mide la cantidad de información de Shannon que un sistema o mecanismo tiene acerca de su estado previo; es decir, los estados de todas sus partes. Tal como yo lo entiendo, la teoría de Tononi presupone un sistema de codificación digital, aunque no necesariamente binario, ya que tiene un repertorio contable de estados de salida.

3. Por ejemplo, el virus del Ébola imita un fragmento celular apoptótico para «ser comido» por un fagocito y así, subirse al tren en una célula que lo transportará por todo el cuerpo (Misasi y Sullivan 2014). Hay muchos ejemplos bien estudiados de camuflaje y mimetismo viral y bacteriano, técnicas copiadas por los biotecnólogos para poder enmascarar nanoartefactos que, en caso contrario, serían atacados por el sistema inmunitario.

4. Supongamos que «dientes más grandes» fuera una mejora de diseño de algún organismo; las materias primas necesarias y la energía para situar estas materias en su posición adecuada no cuentan como información semántica; en cambio, sí cuentan los controles de desarrollo necesarios para lograr esta modificación de diseño.

5. El Ratio Club, fundado en 1949 por el neurólogo John Bates en la Universidad de Cambridge, contaba entre sus miembros a Donald MacKay, Alan Turing, Grey Walter, I. J. Good, William Ross Ashby y Horace Barlow, entre otros. ¡Imaginemos cómo debían ser sus reuniones!

6. La ciencia ha ampliado muchísimo nuestra capacidad de descubrir diferencias que marcan la diferencia. Un niño puede saber la edad de un árbol contando sus anillos de crecimiento, mientras que un biólogo evolutivo puede saber, aproximadamente, cuántos millones de años atrás dos especies de aves compartieron un ancestro común contando las diferencias de sus ADN. Pero estos elementos de información sobre el tiempo no tienen ningún papel en el diseño del árbol o de las aves; esta información no es útil para ellos, pero ahora se ha convertido en una información útil para nosotros.

7. Colgate y Ziock (2010) defienden una definición de información como «lo que se selecciona», que, sin duda, casa bien con mi propia definición,

pero para que pueda encajar en los casos que consideran, el término se selecciona tiene que divagar un poco.

8. Por desgracia, Gibson no solo ignora la cuestión de qué maquinaria interna logra esta captación; a menudo parece negar que en este tema haya ninguna pregunta difícil a responder. El lema de los gibsonianos radicales es un buen resumen de esta postura, que yo no comparto: «no se trata de lo que hay en tu cabeza, sino en lo que está tu cabeza».

9. Debo agradecer a Kim Sterelny y a David Haig haberme hecho notar estos resultados.

10. Puede surgir información útil incluso a escala molecular. David Haig, en su fascinante ensayo «The Social Gene» (1997) aprovecha la perspectiva agencial hasta llegar al concepto de lo que él denomina el gen estratégico. Tal como señala, «el origen de las moléculas capaces de diferenciar entre ellas mismas y moléculas muy parecidas multiplicó la cantidad de estrategias disponibles para los genes y permitió la evolución de grandes agregados multicelulares» (p. 294). Cuando no hay información disponible para explotar, los genes no pueden rendirse ni formar coaliciones con mejores opciones; una anticipación de la madre naturaleza del principio del pelotón de fusilamiento.

11. Debo agradecer a Ron Planer esta sugerencia.

12. La definición de Colgate y Ziock (2010) tiene una condición adicional que rechazo explícitamente: «seleccionar información de un modo útil, información que debe almacenarse (escribirse); en caso contrario, no hay forma de decidir qué es lo que se ha seleccionado» (p. 58). Todo depende de lo que signifique «almacenarse (escribirse)». Yo diría que la información acerca de la construcción de nidos se almacena y se transmite en la estirpe del pájaro, pero no está anotada (como información sobre construcción de nidos). Tal como me recordó Paul Oppenheim (correspondencia personal), el clásico de F. C. Bartlett Remembering (1932) avisaba frente al hecho de considerar los recuerdos como la recuperación de algo que se había almacenado en algún lugar («la memoria») del cerebro.

13. Robert Mathai señala (correspondencia personal) que el desaprendizaje evolutivo nunca es jetsam, sino flotsam que a lo largo del tiempo acaba pareciendo jetsam. No fue lanzado por la borda a

conciencia, pero su accidental caída al agua ha resultado ser una buena cosa, que ha salvado la embarcación.

14. Acuñé el término Vasto (con uve mayúscula) como abreviación de «Verdaderamente mucho más que ASTronómicO», para referirme a cantidades finitas pero inimaginablemente más grandes que las cantidades meramente astronómicas, como el número de microsegundos desde el Big Bang multiplicado por el número de electrones en el universo visible (Dennett 1995). La Biblioteca de Babel es finita pero Vasta. Su opuesto es Vastamente pequeño.

15. He observado que muchas personas se resisten a la motivación flotante, bastante evidente, de por qué son las hembras las que evalúan y los machos los que se pavonean y se embarcan en una señalización costosa: la inversión parental asimétrica. En las especies en que las hembras suelen invertir más tiempo y energía por descendencia que los machos (creando huevos y no esperma, lactando, incubando, criando, etc.) estas deben ser exigentes al aceptar una pareja. Una hembra solo puede hacer una cantidad limitada de huevos u ocuparse de una cantidad limitada de crías, y si escoge aparearse con un macho de segunda clase, puede desperdiciar toda (o casi toda) su capacidad de reproducción, mientras que si un macho escoge aparearse con una hembra de segunda categoría, perderá solo una pequeña parte de su precioso tiempo y podrá hacer más esperma sin problema. En las especies en que la inversión parental es más o menos igual en machos que en hembras, ambos sexos tienen un aspecto similar y actúan del mismo modo.

CAPÍTULO 8

1. La genealogía de los artefactos no es como la de los seres vivos, pero las semejanzas entre la I+D de ambas son sorprendentes, y serán el tema de un capítulo posterior.

2. La distinción entre lenguajes de programación compilados e interpretados solía ocupar un lugar preponderante, pero aquí no nos incumbe y, además, la mayoría de lenguajes actuales son híbridos. En un

lenguaje compilado, el compilador toma como datos de entrada un programa en código fuente y extrae un programa ejecutable en lenguaje de máquina. En un lenguaje interpretado, el intérprete traduce cada instrucción a medida que se escribe (proceso denominado, a veces, traducción dinámica); esto permite al programador probar la ejecución de cada instrucción mientras la escribe, con lo que se facilita considerablemente (a veces) el proceso de desarrollo del programa, a cambio de un cierto coste (a veces pequeño) en la eficiencia final del programa. El Lisp era, en su origen, un lenguaje interpretado.

3. Es bastante divertido señalar que una cosa que las computadoras ajedrecistas no hacen bien es abandonar en partidas sin esperanza. Sus «operadores» humanos suelen poder intervenir en estos casos desesperados, ya que el programa no tiene sentimientos de vergüenza.

4. Hay estadísticos antibayesianos preocupados por su dependencia de un punto de partida «subjetivo»: teniendo en cuenta tus experiencias hasta la fecha, ¿a qué conclusiones deberías llegar en esta ocasión? ¿De dónde sacas tus probabilidades anteriores? ¿Qué pasa si las mías son diferentes? Gelman 2008 es un repaso claro y accesible de esta y otras objeciones. En cualquier caso, esta objeción es irrelevante para nuestra aplicación del pensamiento bayesiano, que está diseñado específicamente para abordar el dilema de un agente bendecido (o maldecido) con un conjunto de probabilidades anteriores que se pregunta qué debe hacer a continuación. La subjetividad del agente es una parte integrante del problema de diseño al que nos enfrentamos.

5. También hay escépticos. Domingos (2015) señala que la puesta en práctica de la generación de expectativas bayesiana depende mucho del algoritmo MCMC (método de Montecarlo con cadenas de Márkov), un algoritmo de fuerza bruta que hace el trabajo duro; culmina las tareas, pero «no simula ningún proceso real» (p. 164). «Los bayesianos, por su parte, seguramente tienen que agradecer al MCMC la creciente popularidad de sus métodos más que a cualquier otra cosa» (p. 165). Quizá sí, pero como los modelos simulan arquitecturas de una estructura masiva en paralelo, su uso del MCMC es casi obligado.

6. El término código heredado (en inglés, legacy code) tiene varios significados; en algunos casos se refiere a código compilado pero que

está restringido a aplicaciones especiales (por ejemplo, para permitir que el programa se ejecute en una plataforma antigua).

7. Una de las virtudes del modelo detallado propuesto por Eliasmith en su pionero trabajo How to Build a Brain (2013) es que su «arquitectura de punteros semánticos» explica la comprensión conductual (percepción y control) del tipo animal, a la vez que deja gran cantidad de plasticidad (palancas a girar) para que los invasores (palabras y otros memes) puedan explotar. No es un modelo de inteligencia «lenguajecéntrico» de inicio. Los detalles de cómo puede funcionar todo esto darían para otro libro entero.

CAPÍTULO 9

1. El hecho de que un diagrama de este tipo esté, en cierto modo, cabeza abajo, con los últimos descendientes más arriba en las ramas que sus ancestros, no es ninguna de nuestras extrañas inversiones, pero si os molesta, poned el diagrama de lado. Nunca he visto un cladograma con el tiempo indicado de derecha a izquierda, pero quizá se usan en textos en árabe o hebreo.

2. Para respetar una distinción técnica que a veces es importante, debemos señalar que en la frase del tipo «el gato duerme en el felpudo» hay dos ocurrencias del tipo el y seis ocurrencias del tipo e. Como los tipos, las ocurrencias son abstracciones, no entidades concretas. Una distinción similar debe hacerse al hablar de genes (aunque a menudo se ignora, con la confusión resultante). En primera aproximación hay por lo menos un caso de cada uno de tus tipos de gen en cada una de tus células, pero también varias ocurrencias de la misma secuencia de codones en tu genoma. Por ejemplo, la duplicación de genes, que tiene un papel fundamental en muchas de las innovaciones más relevantes de la evolución, es la amplificación de una única copia múltiple de un caso de un gen en billones de casos duplicados en las células de la descendencia. Para una clara descripción, véase Dawkins (2004).

3. Para imaginar este fenómeno, es de ayuda compararlo con Shazam, la aplicación para móviles que permite identificar fragmentos de música a través del micrófono. La señal transducida por el micrófono no tiene que «reconvertirse en sonidos» para que el programa la identifique.

4. Sabemos que hay regiones del cerebro en que se almacena el conocimiento de diferentes clases de palabras: para la comida, los instrumentos musicales, las herramientas… Si estas regiones quedan afectadas por un ictus u otras complicaciones cerebrales, se produce una pérdida selectiva de conocimiento de los significados. Así, tus casos de violín y mis casos de violín pueden involucrar actividad en la misma zona del cerebro, pero sería mucho pedir que haber localizado violín, trompeta y vihuela en mi cerebro ayudara a los investigadores a localizarlos en el tuyo.

5. Las connotaciones problemáticas del término cotidiano palabra han llevado a Jackendoff a proponer y definir un sustituto con más precisión técnica: elemento léxico. Por ejemplo, no todos los elementos léxicos tienen pronunciación.

6. Los homínidos son la categoría de primates que incluye a los chimpancés y otros grandes simios, mientras que la categoría más reducida de los homininos es la rama de seis millones de años de antigüedad que solo incluye a los humanos y a sus parientes y ancestros más cercanos.

7. Los simbiontes se clasifican en tres tipos: parásitos, que son perjudiciales para la aptitud genética del anfitrión; comensales, que son neutros (pero «se sientan a la misma mesa»), y mutualistas, que benefician al anfitrión y mejoran su aptitud genética.

8. Aquí me refiero a la información semántica en su sentido más amplio, elaborado en el capítulo 6, la información (semántica) sobre la sintaxis, la fonología (en el caso de bebés que oyen), la pragmática, los contextos y los significados (semántica) de los elementos que se adquieren.

9. Deacon (1997) ofrece una buena descripción, Véase también Deacon (2003) y Dennett (2003b), ambos en Depew y Weber (ed.), Evolution and Learning: The Baldwin Effect Reconsidered. Más recientemente, Christiansen y Chater (2008) presentan una ambiciosa defensa de una versión exagerada de esta tesis en un artículo destacado de la revista

Behavioral and Brain Sciences, en que estudian las alternativas y elaboran una colección especialmente valiosa de respuestas.

10. He aquí un caso en que el diseño inteligente de arriba abajo de la GOFAI resulta particularmente vívido: si consideramos el conocimiento de un agente como un conjunto de axiomas y teoremas, como la geometría euclidiana, necesitaremos «definir nuestros términos». El hecho de que el proyecto Cyc tenga tantos miles de definiciones codificadas a mano en su base de datos es un testimonio de la convicción que algunos tienen en que la IA de arriba abajo es el camino que debe seguirse.

11. Un Buen Truco puede ser «obvio» para la selección natural pero indescifrable para los investigadores humanos curiosos. No debe olvidarse nunca la Segunda Regla de Orgel: la evolución es más lista que tú.

12. Véase Cloud (2015) para una presentación de este tema. La descripción de Cloud es una corrección amable y significativa a mis observaciones anteriores sobre la domesticación de las palabras, no una refutación; aun así, todavía hay algunos desacuerdos menores entre nosotros que tendremos que resolver más adelante.

13. Quizá esto sea un poco exagerado. Hay muchas lenguas en las que no hay ninguna palabra para palabra y puede que sus hablantes no se hayan dado cuenta de que sus lenguas se pueden dividir en partes. Del mismo modo que a los niños se les tiene que enseñar cómo «suenan» las palabras cuando aprenden a deletrear, todo el mundo tiene que aprender que lo que hablan de forma tan natural es una cosa formada por partes reutilizables. Esta parte de la lingüística popular puede que no sea obvia para todo el mundo como lo es para nosotros, verbívoros conscientes.

CAPÍTULO 10

1. A menudo, esta suposición formará parte de una motivación flotante, no algo que el receptor suponga conscientemente. Un niño crédulo puede

ser víctima de unos embusteros sin haber desarrollado todavía una apreciación de lo que es la verdad y la mentira.

2. Durante años, esta hostilidad fue alentada por Stephen Jay Gould, cuya campaña contra el «fundamentalismo darwinista» en general y contra Dawkins en particular convenció a muchos no biólogos de que deberían aliarse con la versión del darwinismo supuestamente más ecuménica y amable con las humanidades que representaba Gould, cuando en realidad, esta versión pagaba un alto precio en confusión y ofuscación a cambio de su halo de amabilidad. El propio Gould afirmaba que solo comprendía vagamente cómo encaja la cultura humana en el mundo en evolución; si queréis entenderlo con más claridad, abandonad a Gould y emprended el camino abierto por Dawkins.

3. La descripción evolutiva que defiende Wilson no es la evolución memética, sino su teoría de «selección grupal multinivel». Tal como he señalado (2006), las condiciones necesarias para que se dé su selección grupal son demasiado poco frecuentes y, además, requieren demasiadas «generaciones» (de grupos, no de individuos) para que resulten eficaces en tiempos históricos. Como, en cualquier caso, Wilson necesita los memes para su descripción, no hay razón para no aceptar la evolución memética como el fenómeno adecuado para explicar la evolución darwiniana inconsciente de elementos funcionales de los grupos sociales.

4. Por error, pensé que Alain escribía acerca de las canoas polinesias cuando cité este fragmento en Dennett 2003. Roger DePledge señaló que Alain (pseudónimo de Émile-Auguste Chartier) estaba hablando «no de pescadores polinesios, sino bretones, de la isla de Groix (47° 38’ N; 3° 28’ O), frente a Lorient».

5. La historia de la ingeniería y las invenciones está llena de ejemplos de malas teorías que bloquean el paso hacia la innovación durante décadas o incluso siglos. Por ejemplo, si te «parece lógico» que no puedes navegar contra el viento, no te molestarás en investigar diseños que podrían lograr tal «imposibilidad».

6. Richerson y Boyd (2005) remarcan que «el modelo de elección racional es un caso límite de la evolución cultural» (p. 175).

7. Está claro que actualmente ningún individuo necesita albergar todos estos memes necesarios para la vida humana, pues desde hace mucho

tiempo hemos establecido una división del trabajo y de las pericias que permite que los individuos se especialicen. Robinson Crusoe es un fascinante y multiforme experimento mental, que demuestra cuán dependientes somos (ya en el siglo XVIII) de gran cantidad de competencias que pocos pueden afirmar que dominan.

8. Por ejemplo, Wilson insiste en que «los detalles [la cursiva es mía] del calvinismo se explican mejor como una adaptación grupal que mediante cualquier otra hipótesis competidora» (p. 119) pero solo unos pocos de estos detalles guardan relación con la solidaridad grupal.

9. O, con la ocurrencia de Richerson y Boyd, «si quieres mejorar la aptitud genética de tus hijos, ¡no les ayudes a hacer los deberes, por el amor de Dios!» (p. 178).

10. Hay muchos relatos de perros que arriesgan su vida (o incluso la dan) por sus dueños, lo que constituye una excepción importante a esta generalización. Los perros se parecen mucho más a los humanos que cualquier otra especie de animal domesticado (incluyendo los gatos) y no es ninguna casualidad; a lo largo de miles de años los humanos los hemos ido seleccionando de forma inconsciente para que muestren esta semejanza. Irónicamente, otro hecho relevante de Homo sapiens y de Canis familiaris es que, a pesar de su habilidad común de subordinar la procreación en aras de una necesidad más elevada, sus poblaciones son mucho más grandes que las de sus parientes más cercanos no domesticados.

11. «La tecnología es un regalo de Dios. Después del regalo de la vida, es quizá el mayor de los regalos divinos. Es la madre de la civilización, de las artes y de las ciencias». Freeman Dyson (1988); al menos Dyson ve nuestro talento como un producto de la tecnología, no como su origen definitivo.

CAPÍTULO 11

1. Un sistema así tendría una interesante propiedad digital: aunque los lectores-escritores deberían poder comprender los textos (las cartas de amor y las amenazas de muerte) usados como dígitos binarios, podrían ser completamente ignorantes del mensaje transmitido en código binario; el mensaje podría estar cifrado o ser un texto simple en sueco, por ejemplo.

2. La localidad de Reading, en Pensilvania, tiene una calle denominada Skyline Drive que sigue una cresta desde la que se divisa la ciudad, un resto fósil del significado antiguo. También podemos pensar en todas las poblaciones cuyo nombre acaba en -ford ('vado' en inglés) aunque ya no haya ningún vado para cruzar el río.

3. Richard (próxima publicación) muestra que el ataque de W. V. O. Quine (1951) contra la famosa distinción entre proposiciones analíticas y sintéticas en la filosofía del lenguaje se puede ver de manera positiva como una ampliación a la filosofía del antiesencialismo de Darwin. Tal como rogaba Ernst Mayr, el pensamiento demográfico debería sustituir a las rígidas taxonomías de todo tipo.

4. Susan Sontag, en su obra Sobre la fotografía (1977), examina el enorme avance epistemológico (comparable a los microscopios y telescopios) que para la ciencia ha representado la fotografía de alta velocidad. «Congelar» el tiempo es tan valioso como ampliar el espacio. El desarrollo reciente de binoculares con estabilización de imagen proporciona las mismas ventajas: el nombre de un barco, ilegible porque no para de moverse en el campo del binocular, ahora se puede leer con facilidad al quedar brevemente congelado.

CAPÍTULO 12

1. Algunos de mis estudiantes se ofenden con esta generalización complaciente e insisten en que no podemos saber si los delfines, por ejemplo, no tienen lenguajes productivos como el nuestro. Lo que puedo admitir es esto: sí, los delfines puede que tengan un lenguaje y una inteligencia a la par de la nuestra y, si es así, han logrado una hazaña

brillante y heroica de camuflaje, ocultándonos estas capacidades y sacrificándose en incontables ocasiones cuando un poco de charla delfínica podría haber salvado sus vidas o las de sus colegas (unos pocos experimentos sugieren que los delfines tienen una limitada capacidad de informar a los demás acerca de ofrecimientos fuera del campo visual).

2. Ya hemos comentado que Richerson y Boyd prefieren no usar el término meme por sus connotaciones de «entidad pseudogenética transmitida de forma discreta y fiel» (p. 63), pero admiten la «necesidad de algún acuerdo oportuno» para lo que denominan «la información almacenada en el cerebro de las personas». Como dejamos que los memes varíen a lo largo de estas dimensiones (veánse las pp. 137-140 de este mismo libro), no nos posicionamos en estas controversias, de modo que, aceptando su resistencia, me apoyo en la definición del Oxford English Dictionary y reformulo sus puntos en el lenguaje de los memes. Un ejercicio útil sería repasar todo el libro Not by Genes Alone y reformular todas sus afirmaciones en términos de memes, para descubrir qué desacuerdos persisten. Por mi parte, los problemas relevantes concernientes a los memes se han refutado en el capítulo 11.

3. «Los ácidos nucleicos inventaron los seres humanos para poder reproducirse incluso en la Luna» (Sol Sipegelman, citado en Eigen 1992).

4. Algunas de las curiosas complejidades del sistema de copia del ADN son producto de miles de millones de años de guerra: la carrera armamentística entre el ADN egoísta, que solo «quiere» ser copiado sin contribuir para nada al mantenimiento y supervivencia del organismo y los sistemas de vigilancia que han ido evolucionando para evitar que este ADN poco cooperativo infeste todo el sistema genético. En la obra de Burt y Tivers Genes in Conflict: The Biology of Selfish Genetic Elements (2008) se puede hallar un fascinante análisis de este tema.

5. «Copiar ciegamente a la mayoría» es una estrategia muy eficaz. Los individuos con competencia (o comprensión conductual) pierden rápidamente su ventaja frente a los copiadores (Rendell et al. 2010).

6. Richerson y Boyd señalan que algunos experimentos con personas y palomas (Kameda y Nakanishi 2002, por ejemplo) muestran la misma división del trabajo sin comunicación entre los participantes.

7. Blackmore (1999) elabora la conjetura de que el lenguaje puede haber empezado como memes parasitarios que, posteriormente, se volvieron mutualistas.

8. ¿Las llamadas de los cercopitecos son innatas o aprendidas? Las llamadas de poblaciones aisladas de cercopitecos son muy similares, lo que sugiere que tales llamadas están fijadas de forma innata, pero los adultos muestran más selectividad que las crías en sus llamadas, que, además, reciben atención con más frecuencia. Así, como siempre, hay una gradación entre comportamiento instintivo y comportamiento aprendido, no una dicotomía (Seyfarth, Cheney y Marler 1980; Cheney y Seyfarth 1990).

9. Hurford tampoco tiene reparos en adoptar un lenguaje intencional para caracterizar las motivaciones (flotantes, en general) de los comportamientos de animales no humanos y de bebés. Por ejemplo: «el perro que muestra los dientes puede ver que cuando lo hace, el otro perro o bien cede o adopta una actitud sumisa […] una señal que ambos entienden tácitamente […] No estamos atribuyendo ningún cálculo consciente a los animales, aunque esta forma "antropomorfizada" de describir la ritualización resulta conveniente» (p. 41).

10. Entre los exploradores que han hecho valiosas incursiones en esta terra incognita, los mejores que empiezan con lo que sabemos acerca del lenguaje y el pensamiento son Jackendoff (2002, 2007, 2007b, 2012) y Millikan (1984, 1993, 2000, 2000b, 2002, 2004, 2005 y próxima publicación).

11. ¿Se trata de una pregunta consciente? No necesariamente; puede ser una mera hambre epistémica, de la variedad flotante que «motiva» el comportamiento exploratorio en todos los animales.

12. La viabilidad de un proceso así se conoce desde la invención del «análisis semántico latente» de Landauer y Dumais (1998), un predecesor de los algoritmos de «aprendizaje profundo» que se aplican actualmente en el Watson de IBM, en Google Translate y en otras aplicaciones espectaculares (véase el capítulo 15).

13. Véase también Christiansen y Chater, «Language as Shaped by the Brain» (artículo destacado en Behavioral and Brain Sciences, 2008) y los extensos comentarios subsiguientes, que constituyen un enérgico debate

acerca de si es la evolución genética o la cultural la que ha tenido un papel más destacado en la creación de nuestras competencias lingüísticas. Christiansen y Chater defienden una postura muy acorde con la mía, pero en algunos aspectos interpretan erróneamente el enfoque memético (véase Blackmore 2008 para consultar más detalles) y exageran los argumentos contra un papel importante de la evolución genética. A estas alturas ya debería estar claro que mi veredicto es que el hardware (estructuras cerebrales transmitidas genéticamente) sigue al software (estructuras cerebrales transmitidas culturalmente) pero actualmente no hay modo de determinar las proporciones relativas. Reformulando a Gibson, podemos decir que la información está en el sonido, pero tiene que haber un equipo en la cabeza para captarlo, y todavía no sabemos cuánta información lleva incorporada ese equipo.

CAPÍTULO 13

1. Vale la pena recordar la posibilidad de que los chimpancés y los bonobos actuales hayan perdido competencias cognitivas o curiosidad durante los seis millones de años que los han separado de nuestro ancestro común. Los peces cavernícolas cuyos ancestros podían ver son ahora ciegos. En un caso más próximo, cuando se hundió el brazo de tierra que unía Australia y Tasmania, hace unos 10.000 años, la población que quedó aislada no solo disminuyó, sino que perdió buena parte de la tecnología de que había disfrutado durante milenios, incluyendo arcos y flechas, barcas y quizá también el dominio del fuego. Véase Diamond (1978), Henrich (2004, 2015) y Ridley (2010), que exponen detalles fascinantes.

2. Esta frase no deja de ser algo parecido a una tautología: la vida surgió y se extinguió, surgió y se extinguió, surgió y se extinguió, hasta que, finalmente, no se extinguió porque consiguió suficientes cosas adecuadas para evitar extinguirse. No es muy informativo, pero tampoco trivial, ya que es una alternativa coherente e incluso confirmable (en principio) a cualquier gancho celeste como teoría milagrosa del origen de la vida.

3. La «entrada encubierta» de información tiene un papel destacado en el humor. Véase Hurley, Dennett y Adams 2011.

4. El libro de Azzouni presenta muchos puntos de vista compatibles con los de este libro (por ejemplo, su discusión de las palabras como herramientas, pp. 83 ss.) y unos pocos, como su rechazo de la distinción entre tipo y caso, que se puede traer a colación con las discusiones aquí presentadas, con ayuda de algunas de sus otras innovaciones, como su separación de la cuantificación y del compromiso ontológico (véanse sus diversos interludios metodológicos).

5. Sperber y Wilson observan que hay un papel que un modelo de competencia podría representar. De acuerdo con su versión, comentan:

Nuestra afirmación es que todos los seres humanos se dirigen automáticamente al procesamiento de información más eficiente posible. Esto es así tanto si son conscientes de ello como si no; de hecho, los diversos y cambiantes intereses conscientes de los individuos proceden de la búsqueda de este objetivo permanente en condiciones cambiantes. (p. 49)

Azzouni señala que Sperber está planteando un modelo de optimización que «se enfrenta a un problema de comprobación. Porque, en general, cualquier proceso cognitivo (sometido a la presión de factores evolutivos) se parecerá a una estrategia de optimización por lo que respecta a los costes del procesamiento frente al valor de lo que se obtiene mediante ese procesamiento» (p. 109).

6. Mientras estaba acabando este libro, recibí un ejemplar de The Centered Mind de Peter Carruthers (2015) y con una rápida ojeada vi que tenía un montón de buenas ideas que casaban con las mías e incluso las anticipaban, acerca de cómo este talento humano de autocuestionarse nos aporta unos poderes cognitivos únicos. Valorar adecuadamente sus contribuciones queda fuera del alcance de este libro, pero es una tarea pendiente. En el anexo, enumero otros libros recientes relevantes que he encontrado demasiado tarde como para incluirlos en mi obra.

7. Experimentos recientes ponen de manifiesto algunas capacidades rudimentarias para la metarrepresentación en aves y mamíferos. En concreto, los animales a los que se les ofrece la oportunidad de escoger entre una opción de alto riesgo y gran recompensa y otra opción más segura mostraron su preferencia por la segunda en los casos en que se les había dado pie a desconfiar de sus intuiciones (Crystal y Foote 2009; otros ensayos en Cook y Weisman 2009).

8. Recordemos aquí la preciosa imagen de Platón en el Teeteto del conocimiento propio como una pajarera llena de aves; estos pájaros son tuyos, pero ¿vienen cuando les llamas?

9. En estos casos, hay margen para el debate y la especulación, así como cierto fundamento histórico para asignar papeles destacados a algunos individuos, no tanto por haber inventado como por haber perfeccionado y propagado el meme, ya sea por decreto (en el caso de gobernantes que establecen la moneda local o el calendario) o aprovechando su influencia (en el caso de los sistemas de escritura o de notación matemática, por ejemplo). Como ejemplo reciente, Al Gore no «inventó Internet», y nunca afirmó haberlo hecho, pero merece el crédito que recibe por su premonitorio apoyo y popularización.

10. Cuando un estudiante me pidió un ejemplo de basura cultural infecciosa difícil de erradicar, le respondí: «Bueno, o sea, es como, o sea, una frase o una expresión que, o sea, no sirve de mucho, pero, o sea, sigues usándola», a lo que me contestó: «Vale, o sea, ya entiendo de qué va, pero yo quería, o sea, un ejemplo».

11. Los escritores han recuperado algunos «segundones» en obras memorables: el poema de Robert Browning sobre Andrea del Sarto (un famoso pintor «perfecto» que vivió a la sombra de Rafael y Miguel Ángel) o la obra Amadeus de Peter Shaffer, una ficción sobre las tribulaciones de Antonio Salieri, eclipsado por el joven Mozart.

12. Los trabajos sobre la normatividad y la aportación de razones procedentes de Sellars en Pittsburgh, a través de Brandom, McDowell y Haugeland, nunca han remarcado, a mi entender, que estas prácticas humanas fundamentales son generadores sistemáticos de ideología falsa siempre que la demanda de razones supera la oferta disponible.

CAPÍTULO 14

1. El periodista científico danés Tor Nørretranders publicó su libro Mærk verden (‘Sentir el mundo’) en 1991, el mismo año en que apareció el mío La conciencia explicada, con mi descripción de la conciencia como una ilusión del usuario. Ninguno de los dos estábamos en situación de podernos citar, lógicamente. La traducción al inglés apareció en 1999 como The User Illusion: Cutting Consciousness Down to Size. Como hice notar en mi libro (nota 9), entre los que propusieron ideas precursoras de esta están Kosslyn (1980), Minsky (1985) y Edelman (1989).

2. Aquí es donde entran en juego los trabajos teóricos y experimentales sobre la imaginería mental de Roger Shepard, Stephen Kosslyn, Zenon Pylyshyn y muchos otros.

3. El operacionalismo es la propuesta de algunos positivistas lógicos de la década de 1920 que afirma que no sabemos lo que significa un término a no ser que podamos definir una operación que podamos usar para determinar cuándo se aplica a algo. Algunos han declarado que el test de Turing se puede considerar una definición operacionalista de inteligencia. El «juego de manos operacionalista» contra el que alerta Searle es la afirmación de que realmente no podemos afirmar saber qué es la conciencia hasta que determinemos cómo podemos aprender acerca de la conciencia de otros. La alternativa de Searle es, en sí misma, un ejemplo muy claro de operacionalismo: si quiero saber qué es la conciencia, mi operación de medida es simple: ¡observo mi interior y la conciencia será lo que sea que vea! A él le sirve, pero no a otros.

CAPÍTULO 15

1. La explicación de Simon Singh, «The Whole Story» (http://simonsingh.net/books/fermats-last-theorem/the-whole-story/) es,

de lejos, la más accessible de cuantas he leído, una versión editada de un ensayo publicado en la revista Prometheus.

2. Watson, especialmente, dispone de algunas capas especializadas de autoobservacion; debe valorar su «confianza» en cada posible respuesta candidata y también puede ajustar su umbral de confianza, arriesgando más o menos en las respuestas. No quiero subestimar a Watson; es un sistema con muchas capas y muchos talentos.

3. La película de ciencia ficción Her de Spike Jonze (2013), con Joaquin Phoenix y la voz de Scarlett Johansson como el personaje virtual, al estilo de Siri, de su teléfono móvil, de quien se enamora, es una de las dos mejores especulaciones sobre el tema, junto con Ex Machina de Alex Garland (2015).

4. Seabright hace hincapié en que ningún grupo de chimpancés o de bonobos puede tolerar la compañía de desconocidos (la proximidad a conespecíficos que no son miembros de la familia o del grupo), una cosa que nosotros experimentamos con serenidad casi cada día. Esto constituye una diferencia significativa. La relativa tranquilidad con que muchas especies de ungulados se agrupan en torno a un abrevadero no es confianza; es una indiferencia instintiva a especies no depredadoras familiares, más como nuestra actitud ante los árboles y los arbustos que como nuestra actitud ante otros seres humanos. La confianza es un fenómeno cultural, tal como señalé en el capítulo 7.

BIBLIOGRAFÍA

Alain, C., Propos d'un Normand 1906-1914, Gallimard, París, 1908 (1956).

Allen, C., M. Bekoff y G. Lauder (ed.), Nature's Purposes: Analyses of Function and Design in Biology, MIT Press, Cambridge, 1998.

Anscombe, G. E. M., Intention, Blackwell, Oxford, 1957.

Arnold, F., 2013. Frances Arnold Research Group. http://www.che.caltech.edu/groups/fha/Projects3b.htm.

Asadia, E., M. Gameiro da Silva, C. Henggeler Antunes, L. Dias y L. Glicksman, «Multi-Objective Optimization for Building Retrofit: A Model Using Genetic Algorithm and Artificial Neural Network and an Application», Energy and Buildings, 81 (2014), pp. 444-456.

Avital, E., y E. Jablonka, Animal Traditions: Behavioural Inheritance in Evolution, Cambridge University Press, Cambridge, 2000.

Azzouni, J., Semantic Perception: How the Illusion of a Common Language Arises and Persists, Oxford University Press, Nueva York, 2013.

Bailey, I. E., F. Muth, K. Morgan, S. L. Meddle y S. D. Healy. 2015. «Birds Build Camouflaged Nests», The Auk, 132, 1 (2015), pp. 11-15.

Baron-Cohen, S., The Essential Difference: Male and Female Brains and the Truth about Autism, Basic Books, Nueva York, 2003 (hay trad. esp.: La gran diferencia: cómo son realmente los cerebros de hombres y mujeres, Amat, 2005).

Bateson, G., Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology, Paladin, St Albans (Australia), 1973.

Baum, E. B., What Is Thought?, MIT Press, Cambridge, 2004.

Behe, M., Darwin's Black Box: The Biochemical Challenge to Evolution, Free Press, Nueva York, 1996 (hay trad. esp.: La caja negra de Darwin: el reto de la bioquímica a la evolución, Andrés Bello, 2000).

Bennett, J. F., Rationality: An Essay Towards an Analysis, Routledge, Londres, 1964.

Bennett, J., Linguistic Behaviour, Cambridge University Press, Londres, 1976.

Berger S. L., T. Kouzarides, R. Shiekhattar y A. Shilatifard, «An Operational Definition of Epigenetics», Genes Dev. 23, 7 (2009), pp. 781-783.

Bernstein, L., «Why Don't You Run Upstairs and Write a Nice Gershwin Tune?», en The Joy of Music, pp. 52-62, Simon and Schuster, Nueva York, 1959.

Beverley, R. M., The Darwinian Theory of the Transmutation of Species Examined (publicación anónima «por un licenciado de la Universidad de Cambridge», Nisbet, Londres, 1868 (citado en una reseña, Athenaeum 2102 [feb. 8], p. 217).

Bickerton, D., Adam's Tongue: How Humans Made Language, How Language Made Humans, Hill and Wang, Nueva York, 2009.

—, More Than Nature Needs: Language, Mind, and Evolution, Harvard University Press, Cambridge, 2014.

Blackmore, S., The Meme Machine, Oxford University Press, Nueva York, 1999 (hay trad. esp.: La máquina de los memes, Paidós, Barcelona, 2000).

—, «Memes Shape Brains Shape Memes», Behavioral and Brain Sciences 31 (2008), p. 513.

—, «Dangerous Memes, or What the Pandorans Let Loose», en Cosmos and Culture: Cultural Evolution in a Cosmic Context, ed. por Steven Dick y Mark Lupisella, pp. 297-318, NASA SP-2009-4802, 2010.

Bogdan, R. J., Mindvaults: Sociocultural Grounds for Pretending and Imagining, MIT Press, Cambridge, 2013.

Borges, J. L., Labyrinths: Selected Stories and other Writings, New Directions, Nueva York, 1962 (edición original en cast.: El jardín de los senderos que se bifurcan, 1941).

Bostrom, N, Superintelligence: Paths, Dangers, Strategies, Oxford University Press, Nueva York, 2014.

Boyd, R. y P. J. Richerson, Culture and the Evolutionary Process, University of Chicago Press, Chicago, 1985.

—, The Origin and Evolution of Cultures, Oxford University Press, Oxford, 2005. http://site.ebrary.com/id/10233633.

Boyd, R., P. Richerson y J. Henrich, «The Cultural Niche: Why Social Learning Is Essential for Human Adaptation», PNAS, 108, supl. 2 (2011), pp. 10918-10925.

Brandom, R., Making It Explicit: Reasoning, Representing, and Discursive commitment, Harvard University Press, Cambridge, 1994.

Bray, D., Wetware: A Computer in Every Living Cell, Yale University Press, New Haven, 2009.

Brenowitz, E. A., D. J. Perkel y L. Osterhout, 2010. «Language and Birdsong: Introduction to the Special Issue», Brain and Language, 115, 1 (2010), pp. 1-2.

Brockman, J. (ed.), This Will Make You Smarter, Harper Torchbook, Nueva York, 2011 (hay trad. esp.: Este libro le hará más inteligente: nuevos conceptos científicos para mejorar su pensamiento, Paidós, Barcelona, 2012).

Bromberger, S., «What Are Words? Comments on Kaplan (1990), on Hawthorne and Lepore, and on the Issue», Journal of Philosophy, 108, 9 (201), pp. 486-503.

Burt, A., y R. Trivers, Genes in Conflict: The Biology of Selfish Genetic Elements, Belknap Press, Cambridge, 2008.

Butterworth, G., «The Ontogeny and Phylogeny of Joint Visual Attention», en Natural Theories of Mind, ed. por A. Whiten, Basil Blackwell, Oxford, 1991, pp. 223-232.

Byrne, R. W., y A. Whiten, Machiavellian Intelligence, Clarendon Press, Oxford, 1988.

Carruthers, P., The Centered Mind: What the Science of Working Memory Shows Us about the Nature of Human Thought, Oxford University Press, Oxford, 2015.

Cavalli-Sforza, L. L., y M. W. Feldman, Cultural Transmission and Evolution: A Quantitative Approach, Princeton University Press, Princeton, 1981.

Chalmers, D., The Conscious Mind: In Search of a Fundamental Theory, Oxford University Press, Nueva York, 1996 (hay trad. esp.: La mente consciente: en busca de una teoría fundamental, Gedisa, Barcelona, 1999).

—, «The Singularity: A Philosophical Analysis», Journal of Consciousness Studies, 17, 9-10 (2010), pp. 7-65.

Cheney, D. L., y R. M. Seyfarth, «Vocal Recognition in Free-Ranging Vervet Monkeys», Animal Behaviour, 28, 2 (1980), pp.: 362-367.

—, «Attending to Behaviour versus Attending to Knowledge: Examining Monkeys' Attribution of Mental States», Animal Behaviour, 40, 4 (1990), pp. 742-753.

Chomsky, N., Aspects of the Theory of Syntax, MIT Press, Cambridge, 1965 (hay trad. esp.: Aspectos de la teoría de la sintaxis, Gedisa, Barcelona, 1999).

—, Reflections on Language, Pantheon Books, Nueva York, 1975 (hay trad. esp.: Reflexiones sobre el lenguaje, Planeta, Barcelona, 1985).

—, The Minimalist Program, MIT Press, Cambridge, 1995 (hay trad. esp.: El programa minimalista, Alianza, Madrid, 1999).

—, New Horizons in the Study of Language and Mind, Cambridge University Press, Cambridge, 2000.

—, On Nature and Language. Cambridge University Press, Nueva York, 2000 (2000b).

—, 2014. «Mysteries and Problems», 18 de octubre, 2014, https://www.youtube.com/watch?v=G8G2QUK_1Wg.

Chou, C., S. Kimbrough, J. Sullivan-Fedock, C. J. Woodard y F. H. Murphy, «Using Interactive Evolutionary Computation (IEC) with Validated Surrogate Fitness Functions for Redistricting», presentado en Genetic and Evolutionary Computation, ACM, Filadelfia, 2012.

Christiansen, M. H. y N. Chater, 2008. «Language as Shaped by the Brain», Behavioral and Brain Sciences, 31, 5 (2008), pp. 489-509.

Christiansen, M. H. y S. Kirby, Language Evolution, Oxford University Press, Oxford, 2003.

Churchland, P. M., Scientific Realism and the Plasticity of Mind, Cambridge University Press, Cambridge, 1979.

Claidière N., K. Smith, S. Kirby y J. Fagot, «Cultural Evolution of Systematically Structured Behaviour in a Non-Human Primate», Proceedings of the Royal Society: Biological Sciences, 281, 1797 (2014), doi: 10.1098/rspb.2014.1541.

Clark, A., «Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science», Behavioral and Brain Sciences, 36, 3 (2013), pp. 181-204.

—, Surfing Uncertainty: Prediction, Action, and the Embodied Mind, Oxford University Press, Nueva York, 2015.

Clark, A. y D. Chalmers, «The Extended Mind», Analysis, 58, 1 (1998), pp. 7-19.

Cloud, D., The Domestication of Language: Cultural Evolution and the Uniqueness of the Human Animal, Columbia University Press, Nueva York, 2014.

Colgate, S. A y H. Ziock, «A Definition of Information, the Arrow of Information, and Its Relationship to Life», Complexity, 16, 5 (2011), pp. 54-62.

Cope, D. y D. R. Hofstadter, Virtual Music: Computer Synthesis of Musical Style, MIT Press, Cambridge, 2001.

Coppinger, R. y L. Coppinger, Dogs: A Startling New Understanding of Canine Origin, Behavior, and Evolution, Scribner, Nueva York, 2001 (hay trad. esp.: Perros: una nueva interpretación sobre su origen, comportamiento y evolución, Ateles/KNS, 2004).

Corballis, M. C., «From Mouth to Hand: Gesture, Speech, and the Evolution of Right-Handedness», Behavioral and Brain Sciences, 26, 2 (2003), pp. 199-208.

—, «The Evolution of Language», Annals of the New York Academy of Sciences, 1156, 1 (2009), pp. 19-43.

Coyne, J, «You Don't Have Free Will», Chronicle of Higher Education (18 de marzo de 2012).

Crick, F., The Astonishing Hypothesis: The Scientific Search for the Soul, Scribner, Nueva York, 1994 (hay trad. esp.: La búsqueda científica del alma, Debate, Madrid, 1994).

Cronin, H., The Ant and the Peacock: Altruism and Sexual Selection from Darwin to Today, Cambridge University Press, Cambridge, 1992.

Crystal, J. D. y A. L. Foote, «Metacognition in Animals», Comparative Cognition and Behavior Review 4 (2009), pp. 1-16.

Darwin, C., On the Origin of Species, New York University Press, Washington, 1859 (hay trad. esp.: El origen de las especies, Espasa Calpe, Madrid, 2008).

—, The Various Contrivances by Which Orchids Are Fertilised by Insects, University of Chicago Press, Chicago, 1984 (1862) (hay trad. esp.: La fecundación de las orquídeas, Laetoli, Madrid, 2007).

—, The Variation of Animals and Plants under Domestication, Orange Judd, Nueva York, 1868 (hay trad. esp.: La variación de los animales y las plantas bajo domesticación, CSIC, Madrid).

—, The Descent of Man and Selection in Relation to Sex, J. Murray, Londres, 1871 (hay trad. esp.: El origen del hombre, Crítica, Barcelona, 2009).

Darwin, C. y F. Burkhardt, The Correspondence of Charles Darwin, vol. 10, Cambridge University Press, Cambridge, 1997.

Dawkins, R., The Selfish Gene, Oxford University Press, Oxford, 1976, ed. rev. 1989 (hay trad. esp.: El gen egoísta, Salvat, Barcelona, 2002).

—, The Blind Watchmaker, Norton, Nueva York, 1986 (hay trad. esp.: El relojero ciego, Tusquets, Basrcelona, 2015).

—, The Ancestor's Tale: A Pilgrimage to the Dawn of Evolution, Houghton Mifflin, Boston, 2004 (hay trad. esp.: El cuento del antepasado, Bosch, Barcelona, 2008).

—, «Extended Phenotype—But Not Too Extended. A Reply to Laland, Turner and Jablonka», Biology and Philosophy, 19 (2004b), pp. 377-396.

Dawkins, R. y J. R. Krebs, «Animal Signals: Information or Manipulation», Behavioural Ecology: An Evolutionary Approach, 2 (1978), pp. 282-309.

de Boer, B. y W. Tecumseh Fitch, «Computer Models of Vocal Tract Evolution: An Overview and Critique», Adaptive Behaviour, 18, 1 (2010), pp. 36-47.

Deacon, T. W., The Symbolic Species: The Co-Evolution of Language and the Brain, W. W. Norton, Nueva York, 1997.

—, Incomplete Nature: How Mind Emerged from Matter, W. W. Norton, Nueva York, 2011.

Debner, J. A. y L. L. Jacoby, «Unconscious Perception: Attention, Awareness, and Control», Journal of Experimental Psychology. Learning, Memory, and Cognition, 20, 2 (1994), pp. 304-317.

Defoe, D., The Life and Strange Surprising Adventures of Robinson Crusoe of York, Mariner: As Related by Himself, E. Stock, Londres, 1883 (hay trad. esp.: Robinson Crusoe, Alianza, Madrid, 2016).

Dehaene, S. y L. Naccache, «Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework», COGNITION, 79, 1-2 (2001), pp. 1-37.

Dennett, D. C., Content and Consciousness, Routledge/Kegan Paul/Humanities Press, Londres/Nueva York, 1969 (hay trad. esp.: Contenido y conciencia, Gedisa, 1996).

—, «Intentional Systems», Journal of Philosophy, 68, 4 (1971), pp. 87-106.

—, Brainstorms: Philosophical Essays on Mind and Psychology, Bradford Books, Montgomery, 1978.

—, «Current Issues in Philosophy of Mind», American Philosophical Quarterly, 15, 4 (1978), pp. 249-261 (1978b).

—, «Why Not the Whole Iguana?», Behavioral and Brain Sciences, 1, 1 (1978), pp. 103-104 (1978c).

—, «Beliefs about Beliefs», comentario sobre Premack y Woodruff, Behavioral and Brain Sciences, 1, 4, pp. 568-570 (1978d).

—, «Passing the Buck to Biology», Behavioral and Brain Sciences, 19 (1980).

—, «Three Kinds of Intentional Psychology», en Reduction, Time and Reality, ed. por R. Healey, Cambridge University Press, Cambridge, 1981, pp. 37-60.

—, «How to Study Consciousness Empirically: Or Nothing Comes to Mind», Synthese, 53 (1982), pp. 159-180.

—, «Intentional Systems in Cognitive Ethology: The ‘Panglossian Paradigm’ Defended»; «Taking the Intentional Stance Seriously», Behavioral and Brain Sciences, 6, 3 (1983), pp. 343-390.

—, Elbow Room: The Varieties of Free Will Worth Wanting, MIT Press, Cambridge, 1984.

—, «Cognitive Wheels: The Frame Problem of AI», en Minds, Machines and Evolution, ed. por C. Hookway, Cambridge University Press, Cambridge, 1984, pp. 129-151, (1984b).

—, «A Route to Intelligence: Oversimplify and Self-monitor», disponible en http://ase.tufts.edu/cogstud/papers/oversimplify.pdf, 1984 (1984c)

—, The Intentional Stance, MIT Press, Cambridge, 1987 (hay trad. esp.: La actitud intencional, Gedisa, Barcelona, 1998).

—, «Cognitive Ethology: Hunting for Bargains or a Wild Goose Chase?», en Goals, No-Goals and Own Goals, ed. por D. Noble y A. Montefiore, Oxford University Press, Oxford, 1989, pp. 101-116.

—, Consciousness Explained, Little Brown, Boston, 1991 (hay trad. esp.: La conciencia explicada, Paidós, Barcelona, 1995).

—, «Real Patterns», Journal of Philosophy, 88, 1 (1991), pp. 27-51 (1991b).

—, «Learning and Labeling», comentario sobre A. Clark y A. Karmiloff-Smith, «The Cognizer's Innards», Mind and Language, 8, 4 (1993), pp. 540-547.

—, «Get Real», Philosophical Topics, 22, 1 (1994), pp. 505-568.

—, Darwin's Dangerous Idea, Simon and Schuster, Nueva York, 1995 (hay trad. esp.: La peligrosa idea de Darwin, Galaxia Gutenberg, Barcelona, 1999).

—, «Do Animals Have Beliefs?», en Comparative Approaches to Cognitive Sciences, ed. por Herbert Roitblat y Jean-Arcady Meyer, MIT Press, Cambridge, 1995, pp. 111-118, (1995b).

—, «Hofstadter's Quest: A Tale of Cognitive Pursuit», Complexity, 1, 6 (1995), pp. 9-12 (1995c).

—, «Animal Consciousness—What Matters and Why», Social Research, 62, 3 (1995), pp. 691-710 (1995d).

—, Kinds of Minds: Toward an Understanding of Consciousness, Basic Books, Nueva York, 1996 (hay trad. esp.: Tipos de mentes: hacia una comprensión de la conciencia, Debate, Madrid, 2000).

—, «Darwinian Fundamentalism: An Exchange», New York Review, 64 (1997).

—, Brainchildren: Essays on Designing Minds, MIT Press, Cambridge, 1998.

—, 1998b. «Reply to Nicholas Humphrey, Cave Art, Autism, and the Evolution of the Human Mind», Cambridge Archeological Journal, 8, 2 (1998), pp. 184-85 (1998b).

—, Reseña de Having Thought: Essays in the Metaphysics of Mind, de John Haugeland, Journal of Philosophy, 96, 8 (1999), pp. 430-435.

—, «In Darwin's Wake, Where Am I?», Proceedings and Addresses of the American Philosophical Association, 75, 2 (2001), pp. 11-30.

—, «The Evolution of Culture», The Monist, 84, 3 (2001), pp. 305-324 (2001b).

—, «Are We Explaining Consciousness Yet?», Cognition, 79 (2001), pp. 221-237 (2001c).

—, «Does Your Brain Use the Images in It, and If So, How?», Behavioral and Brain Sciences, 25, 2 (2002), pp. 189-190.

—, «Tarbutniks Rule», Journal of Evolutionary Biology, 15, 2 (2002), pp. 332-334 (2002b).

—, «Explaining the 'Magic' of Consciousness», Exploring Consciousness, Humanities, Natural Science, Religion, Milán, 19-20 de noviembre de 2001, Proceedings of the International Symposium, Fondazione Carlo Erba, Milán, 2002, pp. 47-58 (2002c).

—, Freedom Evolves, Viking, Nueva York, 2003 (hay trad. esp.: La evolución de la libertad, Paidós, Barcelona, 2004).

—, «The Baldwin Effect: A Crane, Not a Skyhook», en Evolution and Learning: The Baldwin Effect Reconsidered, ed. por Bruce H. Weber y David J. Depew, MIT Press, Cambridge, 2003, pp. 60-79 (2003b).

—, «Who's on First? Heterophenomenology Explained», Journal of Consciousness Studies, 10, 9-10 (2003), pp. 19-30 (2003c).

—, Sweet Dreams: Philosophical Obstacles to a Science of Consciousness, Bradford Book/MIT Press, Cambridge, 2005 (hay trad. esp.: Dulces sueños: obstáculos filosóficos para una ciencia de la conciencia, Katz, Buenos Aires, 2006).

—, Breaking the Spell: Religion as a Natural Phenomenon, Viking, Nueva York, 2006 (hay trad. esp.: Romper el hechizo: la religión como un fenómeno natural, Katz, Buenos Aires, 2007).

—, «The Selfish Gene as a Philosophical Essay», en Richard Dawkins: How a Scientist Changed the Way We Think, ed. por A. Grafen y M. Ridley, Oxford University Press, Oxford, 2006, pp. 101-115 (2006b).

—, «From Typo to Thinko: When Evolution Graduated to Semantic Norms», en Evolution and Culture, ed. por S. Levinson y P. Jaisson, MIT Press, Cambridge, 2006, pp. 101-115 (2006c).

—, «Two Steps Closer on Consciousness», en Paul Churchland, ed. por Brian Keeley, Cambridge University Press, Nueva York, 2006, pp. 193-209 (2006d).

—, «Instead of a Review», Artificial Intelligence, 171, 18 (2007), pp 1110-1113.

—, «Heterophenomenology Reconsidered», Phenomenology and the Cognitive Sciences, 6, 1-2 (2007), pp. 247-270 (2007b).

—, «My Body Has a Mind of Its Own», en Distributed Cognition and the Will: Individual Volition and Social Context, ed. por D. Ross, D. Spurrett, H. Kincaid y G. L. Stephens, MIT Press, Cambridge, 2007, pp. 93-100 (2007c).

—, «Competition in the Brain», en What Have You Changed Your Mind About?, ed. por John Brockman, HarperCollins, Nueva York, 2008, pp.

37-42.

—, «Darwin's 'Strange Inversion of Reasoning'», Proceedings of the National Academy of Sciences of the United States of America, 106 (2009), pp. 10061-10065.

—, «What Is It Like to be a Robot?», reseña de Guilty Robots, Happy Dogs, de David McFarland, BioScience, 59, 8 (2009), pp. 707-709 (2009b).

—, «The Evolution of Why?», en Reading Brandom: On Making It Explicit, ed. por B. Weiss y J. Wanderer, Routledge, Nueva York, 2010, pp. 48-62.

—, «The Mystery of David Chalmers», Journal of Consciousness Studies, 19, 1-2 (2012), pp. 86-95.

—, Erasmus: Sometimes a Spin Doctor Is Right, Praemium Erasmianum Foundation, Amsterdam, 2012 (2012b).

—, «Cycles», en This Will Make You Smarter, ed. por J. Brockman y Edge.org, Harper Torchbook, Nueva York, 2012, pp. 110-119 (2012c) (hay trad. esp.: Este libro le hará más inteligente, Paidós, Barcelona, 2012).

—, Intuition Pumps and Other Tools for Thinking, W. W. Norton, Nueva York, 2013 (hay trad. esp.: Bombas de intuición y otras herramientas de pensamiento, FCE, México, 2015).

—, «Aching Voids and Making Voids», reseña de Incomplete Nature: How Mind Emerged from Matter, de Terrence W. Deacon, Quarterly Review of Biology, 88, 4 (2013), pp. 321-324 (2013b).

—, «Expecting Ourselves to Expect: The Bayesian Brain as a Projector», Behavioral and Brain Sciences, 36, 3 (2013), pp. 209-210 (2013c).

—, «Turing's 'Strange Inversion of Reasoning'», en Alan Turing: His Work and Impact, ed. por S. Barry Cooper y J. van Leeuwen, Elsevier, Amsterdam, 2013, pp. 569-573 (2013d).

—, «Daniel Dennett on Free Will Worth Wanting», en Philosophy Bites Again, ed. por D. Edmonds y N. Warburton, Oxford University Press, Nueva York, 2014, pp. 125-133.

—, «Reflections on Free Will», reseña de Free Will, de Sam Harris. Naturalism.org (2014b).

—, «The Singularity—An Urban Legend?», en What to Think about Machines That Think, ed. por John Brockman, HarperCollins, Nueva York, 2015, pp. 85-88.

—, «The Friar's Fringe of Consciousness», en Structures in the Mind: Essays on Language, Music, and Cognition in Honor of Ray Jackendoff, ed. por Ida Toivonen, Piroska Csuri y Emile van der Zee, MIT Press, Cambridge, 2015, pp. 371-378 (2015b).

—, «Why and How Does Consciousness Seem the Way It Seems?», en Open MIND, ed. por T. Metzinger y J. M. Windt, MIND Group, Fráncfort del Meno, 2015, doi: 10.15502/9783958570245 (2015c).

—, «How Our Belief in Qualia Evolved, and Why We Care So Much—A Reply to David H. Baßler» en Open MIND, ed. por T. Metzinger y J. M. Windt, MIND Group, Fráncfort del Meno, 2015, doi: 10.15502/9783958570665 (2015d).

—, «Jonathan Bennett's Rationality», en Ten Neglected Classics, ed. por Eric Schliesser, próxima publicación.

Dennett, D. C. y R. T. McKay, «A Continuum of Mindfulness», Behavioral and Brain Sciences, 29 (2006), pp. 353-354.

Descartes, R., Discourse on Method, Liberal Arts Press, Nueva York, 1956 (1637) (hay trad. esp. del original francés: Discurso del método, Alianza, Madrid, 1999).

—, Méditations métaphysiques, Michel Soly, París, 1641 (hay trad. esp.: Meditaciones metafísicas, Alianza, Madrid, 2009).

Diamond, J., «The Tasmanians: The Longest Isolation, the Simplest Technology», Nature, 273 (1978), pp. 185-186.

Diesendruck, G. y Lori Markson, «Children's Avoidance of Lexical Overlap: A Pragmatic Account», Developmental Psychology, 37, 5 (2001), pp. 630-641.

Domingos, P., The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World, Basic Books, Nueva York, 2015.

Drescher, G. L., Made-up Minds: A Constructivist Approach to Artificial Intelligence, MIT Press, Cambridge, 1991.

Dyson, F. J., Infinite in All Directions, Gifford Lectures Given at Aberdeen, Scotland, April-November 1985, Harper and Row, Nueva York, 1988 (hay trad. esp.: El infinito en todas direcciones, Tusquets, Barcelona, 1991).

Edelman, G. M., The Remembered Present: A Biological Theory of Consciousness, Basic Books, Nueva York, 1989.

Eigen, M., Steps Towards Life, Oxford University Press, Oxford, 1992.

Eldredge, N., «A la recherche du docteur Pangloss», Behavioral and Brain Sciences, 6, 3 (1983), pp. 361-362.

Eliasmith, C., How to Build a Brain: A Neural Architecture for Biological Cognition, Oxford University Press, Nueva York, 2013.

Emery, N. J., «The Eyes Have It: The Neuroethology, Function and Evolution of Social Gaze», Neuroscience & Biobehavioral Reviews, 24 (2000), pp. 581-604.

Everett, D. L., «Coherent Fieldwork», en Linguistics Today, ed. por Piet van Sterkenberg, John Benjamins, Amsterdam, 2004, pp. 141-162.

Fisher, D., «Swimming and Burrowing in Limulus anti Mesolimulus», Fossils and Strata, 4 (1975), pp. 281-290.

Fitch, W. T., «Nano-Intentionality: A Defense of Intrinsic Intentionality», Biology & Philosophy, 23, 2 (2008), pp. 157-177.

—, The Evolution of Language, Cambridge University Press, Cambridge, 2010, http://dx.doi.org/10.1017/CBO9780511817779.

Fitch, W. T., L. Huber y T. Bugnyar, «Social Cognition and the Evolution of Language: Constructing Cognitive Phylogenies», Neuron, 65, 6 (2010), pp. 795-814.

FitzGibbon, C. D. y J. H. Fanshawe, «Stotting in Thomson's Gazelles: An Honest Signal of Condition», Behavioral Ecology and Sociobiology, 23, 2 (1988), pp. 69-74.

Floridi, L., Information: A Very Short Introduction, Oxford University Press, Oxford, 2010.

Fodor, J., «Review of Steven Pinker's How the Mind Works, and Henry Plotkin's Evolution in Mind», London Review of Books. Reimpreso en Fodor, In Critical Condition, Bradford Book/MIT Press, Cambridge, 1998.

—, LOT 2: The Language of Thought Revisited, Clarendon Press, Oxford, 2008.

Francis, R. C., Why Men Won't Ask for Directions: The Seductions of Sociobiology, Princeton University Press, Princeton, 2004.

Frischen, A., A. P. Bayliss y S. P. Tipper, «Gaze cueing of Attention: Visual Attention, Social Cognition, and Individual Differences», Psychological Bulletin, 133, 4 (2007), pp. 694-724.

Friston, K., M. Levin, B. Sengupta y G. Pezzulo, «Knowing One's Place: A Free-Energy Approach to Pattern Regulation», Journal of the Royal Society Interface, 12 (2015), 20141383.

Frith, C. D., «The Role of Metacognition in Human Social Interactions», Philosophical Transactions of the Royal Society B: Biological Sciences, 367, 1599 (2012), pp. 2213-2223.

Gelman, A., «Objections to Bayesian Statistics», Bayesian Anal., 3, 3 (2008), pp. 445-449.

Gibson, J. J., «The Problem of Temporal Order in Stimulation and Perception», Journal of Psychology, 62, 2 (1966), pp. 141-149.

——, 1979. The Ecological Approach to Visual Perception, Houghton Mifflin, Boston, 1979.

Godfrey-Smith, P., «Postscript on the Baldwin Effect and Niche Construction», en Evolution and Learning: The Baldwin Effect Reconsidered, ed. por Bruce H. Weber y David J. Depew, MIT Press, Cambridge, 2003, pp. 210-223.

—, «Conditions for Evolution by Natural Selection», Journal of Philosophy, 104 (2007), pp. 489-516.

—, Darwinian Populations and Natural Selection, Oxford University Press, Oxford, 2009.

Gorniak, P. y D. Roy, «Perceived Affordances as a Substrate for Linguistic Concepts», MIT Media Lab, 2006.

Gould, S. J., Wonderful Life: The Burgess Shale and the Nature of History, W. W. Norton, Nueva York, 1989 (hay trad. esp.: La vida maravillosa: Burgess Shale y la naturaleza de la historia, Crítica, Barcelona, 1991).

—, Bully for Brontosaurus: Reflections in Natural History, W. W. Norton, Nueva York, 1991 (hay trad. esp.: «Brontosaurus» y la nalga del ministro: reflexiones sobre historia natural, Crítica, Barcelona, 1993).

—, «Darwinian Fundamentalism», parte I de la reseña de Darwin's Dangerous Idea, New York Review of Books, 12 de junio de 1997.

—, «Evolution: The Pleasures of Pluralism», parte I de la reseña de Darwin's Dangerous Idea, New York Review of Books, 26 de junio de 1997 (1997b).

Gould, S. J. y R. C. Lewontin, «The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme», Proceedings of the Royal Society of London, the Evolution of Adaptation by Natural Selection (Sept. 21), Series B, Biological Sciences, 205, 1161 (1979), pp. 581-598.

Graziano, M. S. A., Consciousness and the Social Brain, Oxford University Press, Oxford/Nueva York, 2013.

Grice, H. P., «Meaning», The Philosophical Review, 66 (1957), pp. 377-388.

—, «Utterer's Meaning, Sentence Meaning, and Word Meaning», Foundations of Language, 4, 1968. Reimpreso como capítulo 6 en Grice 1989, pp. 117-137.

—, «Utterer's Meaning and Intentions», Philosophical Review, 78, 1969. Reimpreso como capítulo 5 en Grice 1989, pp. 86-116.

—, Intention and Uncertainty, Oxford University Press, Londres, 1972.

—, Studies in the Way of Words. The 1967 William James Lectures at Harvard University, Harvard University Press, Cambridge, 1989.

Griffin, D. R. y C. A. Ristau, «Aspects of the Cognitive Ethology of an Injury-Feigning Bird, the Piping Plover», en Cognitive Ethology: The Minds of Other Animals. Essays In Honor of Donald R. Griffin, L. Erlbaum Associates, Hillsdale, 1991.

Griffin, R. y D. C. Dennett, «What Does The Study of Autism Tell Us about the Craft

of Folk Psychology?», en Social Cognition: Development, Neuroscience, and Autism, ed. por T. Striano y V. Reid, Wiley-Blackwell, Malden, 2008, pp. 254-280.

Griffiths, P., «The Cronin Controversy», Brit. J. Phil. Sci., 46 (1995), pp. 122-138.

—, «Molecular and Developmental Biology», en The Blackwell Guide to the Philosophy of Science, ed. por P. Machamer y M. Silverstein, Blackwell, Oxford, 2008, pp. 252-271.

Guston, P., Philip Guston: Collected Writings, Lectures, and Conversations, ed. por C. Coolidge, University of California Press, Berkeley/Los Ángeles/Londres, 2011.

Haig, D., «The Social Gene», en Behavioural Ecology: An Evolutionary Approach, ed. por J. R. Krebs y N. Davies, Blackwell Science, Oxford, 1997, pp. 284-304.

—, «Conflicting Messages: Genomic Imprinting and Internal Communication», en Sociobiology of Communication: An Interdisciplinary Perspective, ed. por P. D'Ettorre y D. P. Hughes, Oxford University Press, Oxford, 2008, pp. 209-223.

Halitschke, R., J. A. Stenberg, D. Kessler, A. Kessler y I. T. Baldwin, «Shared Signals—'Alarm Calls' from Plants Increase Apparency to Herbivores and Their Enemies in Nature», Ecology Letters, 11, 1 (2008), pp. 24-34.

Hansell, M. H., Bird Nests and Construction Behaviour, Cambridge University Press, Cambridge, 2000.

—, Animal Architecture, Oxford University Press, Oxford, 2005.

—, Built by Animals, Oxford University Press, Oxford, 2007.

Hardy, A., «Was Man More Aquatic in the Past?», The New Scientist, 1960, pp. 642-645.

Hardy, T., Selected Poems of Thomas Hardy, Macmillan, Londres, 1960.

Harris, S., Free Will, Free Press, Nueva York, 2012.

Harvey, I., P. Husbands, D. Cliff, A. Thompson y N. Jakobi, «Evolutionary Robotics: the Sussex Approach», Robotics and Autonomous Systems, 20, 2-4 (1997), pp. 205-224.

Haugeland, J., Artificial Intelligence: The Very Idea, MIT Press, Cambridge, 1985.

—, Having Thought: Essays in the Metaphysics of Mind, Harvard University Press, Cambridge, 1998.

Hauser, M. D., The Evolution of Communication, MIT Press, Cambridge, 1996.

Hauser, M. D., N. Chomsky y W. Tecumseh Fitch, «The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?», Science, 298, 5598 (2002), pp. 1569-1579.

Heeks, R. J., «Discovery Writing and the So-called Forster Quote», 13 de abril de 2011, https://rjheeks.wordpress.com/2011/04/13/discovery-writing-and-the-so-called-forster-quote/.

Henrich, J., «Demography and Cultural Evolution: Why Adaptive Cultural Processes Produced Maladaptive Losses in Tasmania», American Antiquity, 69, 2 (2004), pp. 197-221.

—, 2015. The Secret of Our Success, Princeton University Press, Princeton, 2015.

Hewes, G. W., The Origin of Man, Burgess, Minneapolis, 1973.

Hinton, G. E., «Learning Multiple Layers of Representation», Trends in Cognitive Sciences, 11, 10 (2007), pp. 428-434.

Hofstadter, D., Gödel, Escher, Bach: An Eternal Golden Braid, Basic Books, Nueva York, 1979 (hay trad. esp.: Gödel, Escher, Bach: un eterno y grácil bucle, Tusquets, Barcelona, 1998).

—, «Reflections», en The Mind's I, ed. por D. Hofstadter y D. Dennett, 1981, pp. 403-404.

—, «Can Creativity Be Mechanized?», Scientific American, 247 (1982), pp. 20-29 (hay trad. esp.: «¿Mecanizar la inspiración?», Investigación y ciencia, 74, noviembre de 1982).

—, «Who Shoves Whom Around Inside the Careenium? Or What Is the Meaning of the Word 'I'?», Synthese, 53, 2 (1982), pp. 189-218 (1982b).

—, Metamagical Themas: Questing for the Essence of Mind and Pattern, Basic Books, Nueva York, 1985.

—, I Am a Strange Loop, Basic Books, Nueva York, 2007 (hay trad. esp.: Yo soy un extraño bucle, Tusquets, Barcelona, 2008).

Hofstadter D. y D. Dennett (ed.), The Mind's I: Fantasies and Reflections on Self and Soul, Basic Books/Harvester, Nueva York/Hassocks, 1981.

Hohwy, J., «Attention and Conscious Perception in the Hypothesis Testing Brain», Frontiers in Psychology, 3, 96 (2012), pp. 1-14.

—, The Predictive Mind, Oxford University Press, Nueva York, 2013.

Huebner, B. y D. Dennett, «Banishing 'I' and 'We' from Accounts of Metacognition», respuesta a P. Carruthers, «How We Know Our Own Minds: The Relationship Between Mindreading and Metacognition», Behavioral and Brain Sciences, 32 (2009), pp. 121-182.

Hughlings Jackson, J., «Hughlings Jackson on Aphasia and Kindred Affections of Speech», Brain, 38 (1915), pp. 1-190.

Hume, D., A Treatise of Human Nature, John Noon, Londres, 1739 (hay trad. esp.: Tratado de la naturaleza humana, Tecnos, Madrid, 2005).

Humphrey, N. K., «The Social Function of Intellect», Growing Points in Ethology (1976), pp. 303-317.

—, Soul Searching: Human Nature and Supernatural Belief, Chatto and Windus, Londres, 1995.

—, Leaps of Faith: Science, Miracles, and the Search for Supernatural Consolation, Basic Books, Nueva York, 1996.

—, «Cave Art, Autism, and the Evolution of the Human Mind», Cambridge Archeological Journal, 8, 2 (1998), pp. 184-185.

—, How to Solve the Mind-Body Problem, Imprint Academic, Thorverton, 2000.

—, Seeing Red: A Study in Consciousness, Harvard University Press, Cambridge, 2006.

—, «The Colour Currency of Nature», en Colour for Architecture Today, ed. por T. Porter y B. Mikellides, Taylor and Francis, Londres, 2009, p. 912.

—, Soul Dust: The Magic of Consciousness, Princeton University Press, Princeton, 2011.

Humphrey, N. K. y D. Dennett, «Speaking for Ourselves: An Assessment of Multiple Personality-Disorder», Raritan—A Quarterly Review, 9, 1 (1989), pp. 68-98.

Hurford, J. R., The Origins of Language: A Slim Guide, Oxford University Press, Nueva York, 2014.

Hurley, M. M., D. C. Dennett y R. B. Adams, Inside Jokes: Using Humor to Reverse-Engineer the Mind, MIT Press, Cambridge, 2011.

Jackendoff, R. Patterns in the Mind, Basic Books, Nueva York, 1994.

—, «How Language Helps Us Think», Pragmatics and Cognition, 4, 1 (1996), p. 1-34.

—, Foundations of Language: Brain, Meaning, Grammar, Evolution, Oxford University Press, Nueva York, 2002.

—, Language, Consciousness, Culture: Essays on Mental Structure, MIT Press, Cambridge, 2007.

—, «Linguistics in Cognitive Science: The State of the Art», Linguistic Review, 24 (2007), pp. 347-401 (2007b).

Jackendoff, R., N. Cohn y B. Griffith, A User's Guide to Thought and Meaning, Oxford University Press, Nueva York, 2012.

Jakobi, N., «Evolutionary Robotics and the Radical Envelope-of-Noise Hypothesis», Adaptive Behavior, 6 (1997), pp. 325-367.

Jolly, A., Lemur Behavior: A Madagascar Field Study, University of Chicago Press, Chicago, 1966.

Kameda, T. y D. Nakanishi, «Cost-benefit Analysis of Social/ Cultural Learning in a Nonstationary Uncertain Environment: An Evolutionary Simulation and an Experiment with Human Subjects», Evolution and Human Behavior, 23, 5 (2002), pp. 373-393.

Kaminski, J., «Dogs (Canis familiaris) are Adapted to Receive Human Communication», en Neurobiology of "Umwelt:" How Living Beings Perceive the World, ed. por A. Berthoz e Y. Christen, Springer Verlag, Berlín, 2009, pp. 103-107.

Kaminski, J., J. Brauer, J. Call y M. Tomasello, «Domestic Dogs Are Sensitive to a Human's Perspective», Behaviour, 146 (2009), pp. 979-998.

Kanwisher N., et al., «The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception», Journal of Neuroscience, 17, 11 (1997), pp. 4302-4311.

Kanwisher, N. y D. Dilks, «The Functional Organization of the Ventral Visual Pathway in Humans», en The New Visual Neurosciences, ed. por L. Chalupa y J. Werner, MIT Press, Cambridge, 2013.

Kaplan, D., «Words», Proceedings of the Aristotelian Society, Supplementary Volumes (1990), pp. 93-119.

Katchadourian, R., «The Doomsday Invention: Will Artificial Intelligence Bring Us Utopia or Destruction?», New Yorker, 23 de noviembre de 2015, pp. 64-79.

Kauffman, S., «The Adjacent Possible», Edge.org, 9 de noviembre de 2003, https://edge.org/conversation/stuart_a_kauffman-the-adjacent-possible.

Keller, H., The World I Live In, Century, Nueva York, 1908 (hay trad. esp.: El mundo en el que vivo, Atalanta, Vilaür, 2012).

Kessler, M. A., y B. T. Werner, «Self-Organization of Sorted Patterned Ground», Science, 299, 5605 (2003), pp. 380-383.

Kobayashi, Y. y N. Yamamura, «Evolution of Signal Emission by Non-infested Plants Growing Near Infested Plants to Avoid Future Risk», Journal of Theoretical Biology, 223 (2003), pp. 489-503.

Kosslyn, S. M., «Information Representation in Visual Images», Cognitive Psychology, 7, 3 (1975), pp. 341-370.

—, Image and Mind, Harvard University Press, Cambridge, 1980.

Kosslyn, S. M., et al., «On the Demystification of Mental Imagery», Behavioral and Brain Sciences, 2, 4 (1979), pp. 535-548.

Kosslyn, S. M., et al., «The Neural Foundations of Imagery», Nature Reviews Neuroscience, 2 (2001), pp. 635-642.

Kurzweil, R., The Singularity Is Near: When Humans Transcend Biology,

Viking, Nueva York, 2005.

Laland, K., J. Odling-Smee y M. W. Feldman, «Group Selection: A Niche Construction Perspective», Journal of Consciousness Studies, 7, 1 (2000), pp. 221-225.

Landauer, T. K., P. W. Foltz y D. Laham, «An Introduction to Latent Semantic Analysis», Discourse Processes, 25, 2-3 (1998), pp. 259-84.

Lane, N., The Vital Question: Why Is Life the Way It Is?, Profile, Londres, 2015 (hay trad. esp.: La cuestión vital:¿por qué la vida es como es?, Ariel, Barcelona, 2016).

Levin, M., «Molecular Bioelectricity: How Endogenous Voltage Potentials Control Cell Behavior and Instruct Pattern Regulation in Vivo», Molecular Biology of the Cell, 25 (2014), pp. 3835-3850.

Levine, J., «Materialism and Qualia: The Explanatory Gap», Pacific Philosophical Quarterly, 64 (1983), pp. 354-361.

Levitin, D. J., «Absolute Memory for Musical Pitch: Evidence from the Production of Learned Melodies», Perception & Psychophysics, 56, 4 (1994), pp. 414-423.

Levitin, D. J., y P. R. Cook, «Memory for Musical Tempo: Additional Evidence That Auditory Memory Is Absolute», Perception & Psychophysics, 58, 6 (1996), pp. 927-935.

Lewis, S. M. y C. K. Cratsley, «Flash Signal Evolution, Mate Choice and Predation in Fireflies», Annual Review of Entomology, 53 (2008), pp. 293-321.

Lieberman, M. D., Social: Why Our Brains Are Wired to Connect, Crown, Nueva York, 2013.

Littman, M. L., S. T. Dumais y T. K. Landauer, «Automatic Cross-Language Information Retrieval Using Latent Semantic Indexing», en Cross-Language Information Retrieval, Springer, Nueva York, 1998, pp. 51-62.

Lycan, W. G., Consciousness, MIT Press, Cambridge, 1987.

MacCready, P., «An Ambivalent Luddite at a Technological Feast», Designfax, agosto de 1999.

MacKay, D. M., «Electroencephalogram Potentials Evoked by Accelerated Visual Motion», Nature, 217 (1968), pp. 677-678.

Markkula, G., «Answering Questions about Consciousness by Modeling Perception as Covert Behavior», Frontiers in Psychology, 6 (2015), pp. 803-815.

Marr, D. y T. Poggio, «From Understanding Computation to Understanding Neural Circuitry», Artificial Intelligence Laboratory. A.I. Memo, MIT, Cambridge, 1976.

Marx, K., Carta a Lasalle, Londres, 16 de enero de 1861, en Marx/Engels Collected Works, International Publishers, Nueva York, 1985.

Mayer, G., «Steps toward the Origin of Life», blog de Jerry Coyne, 2009, https://whyevolutionistrue.wordpress.com/2009/05/15/steps-toward-the-origin-of-life/.

McClelland, J. y J. Bybee, «Gradience of Gradience: A Reply to Jackendoff», Linguistic Review, 24 (2007), pp. 437-455.

McCulloch, W. S. y W. Pitts, «A Logical Calculus of the Ideas Imminent in Nervous Activity», Bulletin of Mathematical Biophysics, 5 (1943), pp. 115-133.

McFarland, D., Problems of Animal Behaviour, Longman Scientific and Technical, Harlow, 1989.

—, «Goals, No-Goals and Own Goals», en Goals, No-Goals and Own Goals: A Debate on Goal-Directed and Intentional Behaviour, ed. por A. Montefiore y D. Noble, Unwin Hyman, Londres, 1989, pp. 39-57 (1989b).

McGeer, V., «Autistic Self-awareness», Philosophy, Psychiatry & Psychology, 11 (2004), pp. 235-251.

McGinn, C., The Problem of Consciousness: Essays towards a Resolution, Blackwell, Cambridge, 1991.

—, The Mysterious Flame: Conscious Minds in a Material World, Basic Books, Nueva York, 1999.

McKay, R. T. y D. C. Dennett, «The Evolution of Misbelief», Behavioral and Brain Sciences, 32, 6 (2009), pp. 493.

Mercier, H. y D. Sperber, «Why Do Humans Reason? Arguments for an Argumentative Theory», Behavioral and Brain Sciences, 34 (2011), pp. 57-111.

Merikle, P. M., D. Smilek y J. D. Eastwood, «Perception without Awareness: Perspectives from Cognitive Psychology», Cognition, 79, 1/2 (2001), pp. 115-134.

Miller, G. F., The Mating Mind: How Sexual Choice Shaped the Evolution of Human Nature, Doubleday, Nueva York, 2000.

Miller, G. A., E. Galanter y K. H. Pribram, Plans and the Structure of Behavior, Henry Holt, Nueva York, 1960 (hay trad. esp.: Planes y estructura de la conducta, Debate, Barcelona, 1983).

Miller, M. B. y B. L. Bassler, «Quorum Sensing in Bacteria», Annual Reviews in Microbiology, 55, 1 (2001), pp. 165-199.

Millikan, R. G., Language, Thought, and Other Biological Categories: New Foundations for Realism, MIT Press, Cambridge, 1984.

—, White Queen Psychology and Other Essays for Alice, MIT Press, Cambridge, 1993.

—, On Clear and Confused Ideas: An Essay about Substance Concepts, Cambridge University Press, Cambridge, 2000.

—, «Naturalizing Intentionality», en Philosophy of Mind, Proceedings of the Twentieth World Congress of Philosophy, ed. por B. Elevitch, Philosophy Documentation Center, 2000, vol. 9, pp. 83-90 (2000b).

—, Varieties of Meaning. The 2002 Jean Nicod Lectures, MIT Press, Cambridge, 2004.

—, Language: A Biological Model, Clarendon Press, Oxford, 2005.

—, Unicepts, Language, and Natural Information, próxima publicación.

Minsky, M., The Society of Mind, Simon and Schuster, Nueva York, 1985.

Misasi J. y N. J. Sullivan, «Camouflage and Misdirection: The Full-on Assault of Ebola Virus Disease», Cell, 159, 3 (2014), pp. 477-486.

Moravec, H. P., Mind Children: The Future of Robot and Human Intelligence, Harvard University Press, Cambridge, 1988.

Mordvintsev, A., C. Olah y M. Tyka, «Inceptionism: Going Deeper into Neural Networks», Google Research Blog, consultado el 20 de junio de 2015.

Morgan, E., The Aquatic Ape, Stein and Day, Nueva York, 1982.

—, 1997. The Aquatic Ape Hypothesis, Souvenir Press, Londres, 1997.

Morin, O., How Traditions Live and Die: Foundations for Human Action, Oxford University Press, Nueva York, 2016.

Nagel, T., «What Is It Like to Be a Bat?», Philosophical Review, 83, 4 (1974), pp. 435-450.

Newell, A. 1992. «The Knowledge Level», Artificial Intelligence, 18, 1 (1992), pp. 87-127.

Nimchinsky, E. A., E. Gilissen, J. M. Allman., D. P. Perl, J. M. Erwin y P. R. Hof, «A Neuronalmorphologic Type Unique to Humans and Great Apes», Proc Natl Acad Sci., 96, 9 (1999), pp. 5268-5273.

Nørretranders, T., The User Illusion: Cutting Consciousness Down to Size,

Viking, Nueva York, 1998.

Peirce, C. S., Collected Papers of Charles Sanders Peirce, ed. por Charles Hartshorne y Paul Weiss, Harvard University Press, Cambridge, 1906.

Penrose, R., The Emperor’s New Mind: Concerning Computers, Minds, and the Laws of Physics, Oxford University Press, Oxford, 1989 (hay

trad. esp.: La nueva mente del emperador, Random House, Barcelona, 1991).

Pinker, S., How the Mind Works, W. W. Norton, Nueva York, 1997 (hay trad. esp.: Cómo funciona la mente, Destino, Barcelona, 2004).

—, «Language as an Adaptation to the Cognitive Niche», Studies in the Evolution of Language, 3 (2003), pp. 16-37.

—, «Commentary on Daniel Dennett», Mind, Brain, and Behavior Lecture, Harvard University, 23 de abril de 2009, https://www.youtube.com/watch?v=3H8i5x-jcew.

Pinker, S. y R. Jackendoff, «The Faculty of Language: What's Special about It?», Cognition, 95, 2 (2005), pp. 201-236.

Powner, M. W., B. Gerland y J. D. Sutherland, «Synthesis of Activated Pyrimidine Ribonucleotides in Prebiotically Plausible Conditions», Nature, 459, 7244 (2009), pp. 239-242.

Pugh, G. E., The Biological Origin of Human Values, Basic Books, Nueva York, 1978.

Pylyshyn, Z., «What the Mind's Eye Tells the Mind's Brain: A Critique of Mental Imagery», Psychological Bulletin, 80 (1973), pp. 1-24.

—, «Mental Imagery: In Search of a Theory», Behavioral and Brain Sciences, 25, 2 (2002), pp. 157-182.

Quine, W. V., «Two Dogmas of Empiricism», Philosophical Review, 60 (1951), pp. 20-43.

Rehder, M. F., M. B. Schreiner, W. Wolfe, D. Laham, T. K. Landauer y W. Kintsch, «Using Latent Semantic Analysis to Assess Knowledge: Some Technical Considerations», Discourse Processes, 25, 2/3 (1998), pp. 337-354.

Rendell L., R. Boyd, D. Cownden, M. Enquist, K. Eriksson, M. W. Feldman, L. Fogarty, S. Ghirlanda, T. Lillicrap y K. N. Laland, «Why Copy Others? Insights from the Social Learning Strategies Tournament», Science, 328, 5975 (2010), pp. 208-213.

Richard, M., Meanings as Species, próxima publicación.

Richerson, P. J. y R. Boyd, Not by Genes Alone, University of Chicago Press, Chicago, 2004.

Ridley, M., The Rational Optimist, Harper Collins, Nueva York, 2010 (hay trad. esp.: El optimista racional, Taurus, Barcelona, 2011).

Ristau, C. A., «Language, Cognition, and Awareness in Animals?», Annals of the New York Academy of Sciences, 406, 1 (1983), pp. 170-186.

Rogers, D. S. y P. R. Ehrlich, «Natural Selection and Cultural Rates of Change», Proceedings of the National Academy of Sciences of the United States of America, 105, 9 (2008), pp. 3416-3420.

Rosenberg, A., The Atheist's Guide to Reality: Enjoying Life without Illusions. W. W. Norton, Nueva York, 2011.

Roy, D., «The Birth of a Word», charla TED, 2011, http://www.ted.com/talks/deb_roy_the_birth_of_a_word.

Sanford, D. H., «Infinity and Vagueness», Philosophical Review, 84, 4 (1975), pp. 520-535.

Scanlon, T., Being Realistic about Reasons, Oxford University Press, Nueva York, 2014.

Schönborn, C., «Finding Design in Nature», New York Times, 7 de julio de 2005.

Schwitzgebel, E., «No Unchallengeable Epistemic Authority, of Any Sort, Regarding Our Own Conscious Experience—Contra Dennett?», Phenomenology and the Cognitive Sciences, 6, 1-2 (2007), pp. 1-2.

Seabright, P., The Company of Strangers: A Natural History of Economic Life, Princeton University Press, Princeton, 2010 (ed. rev.).

Searle, J. R., «Minds, Brains, and Programs», Behavioral and Brain Sciences, 3, 3 (1980), pp. 417-457.

—, The Rediscovery of the Mind, MIT Press, Cambridge, 1992 (hay trad. esp.: El redescubrimiento de la mente, Crítica, Barcelona, 1996).

Selfridge, O. G., «Pandemonium: A Paradigm for Learning in Mechanisation of Thought Processes», en Proceedings of a Symposium Held at the National Physical Laboratory, 1958, pp. 513-526.

Sellars, W., Science, Perception, and Reality, Routledge and Paul, Londres, 1962.

Seung, H. S., «Learning in Spiking Neural Networks by Reinforcement of Stochastic Synaptic Transmission», Neuron, 40, 6 (2003), pp. 1063-1073.

Seyfarth, R. y D. Cheney, «The Assessment by Vervet Monkeys of Their Own and Another Species' Alarm Calls», Animal Behaviour, 40, 4 (1990), pp. 754-764.

Seyfarth, R., D. Cheney y P. Marler, «Vervet Monkey Alarm Calls: Semantic Communication in a Free-Ranging Primate», Animal Behaviour, 28, 4 (1980), pp. 1070-1094.

Shanahan, M., Embodiment and the Inner Life, Oxford University Press, Nueva York, 2010.

Shannon, C. E., «A Mathematical Theory of Communication», Bell System Technical Journal, 27, 3 (1948).

Shannon, C. E. y W. Weaver, The Mathematical Theory of Communication, University of Illinois Press, Urbana, 1949 (hay trad. esp.: Teoría matemática de la comunicación, Forja, Madrid, 1981).

Shepard, R. N., y J. Metzler, «Mental Rotation of Three Dimensional Objects», Science, 171, 3972 (1971), pp. 701-703.

Shepard, R. N., y L. A. Cooper, Mental Images and Their Transformations, MIT Press, Cambridge, 1982.

Siegel, L., Net of Magic: Wonders and Deceptions in India, University of Chicago Press, Chicago, 1991.

Siewert, C., «In Favor of (Plain) Phenomenology», Phenomenology and the Cognitive Sciences, 6, 1-2 (2007), pp. 201-220.

Simmons, K. E. L., «The Nature of the Predator-Reactions of Breeding Birds», Behaviour, 4 (1952), pp. 161-171.

Simon, H. A., The Sciences of the Artificial, MIT Press, Cambridge, 1969.

Skutch, A. F., Parent Birds and Their Young, University of Texas Press, Austin, 1976.

Smith, B. C., «The Limits of Correctness in Computers», simposio sobre guerra nuclear no intencionada, 5.º congreso internacional de médicos para la prevención de la guerra nuclear, Budapest, Hungría, 28 de junio-1 de julio de 1985.

Smith, S. D., y P. M. Merikle, «Assessing the Duration of Memory for Information Perceived without Awareness», póster presentado en la 3.ª reunión anual de la Asociación para el Estudio Científico de la Conciencia, Canadá, 1999.

Smith, S., Not Waving but Drowning; Poems, A. Deutsch, Londres, 1957.

Sober, E., y D. Sloan Wilson, «Some Varieties of Greedy Ethical Reductionism», en Darwin's Dangerous Idea, 1995, pp. 467-481.

Sontag, S., On Photography, Farrar, Straus and Giroux, Nueva York, 1977 (hay trad. esp.: Sobre la fotografía, EDHASA, Barcelona, 1996).

Specter, M., «The Gene Hackers: The Promise of CRISPR Technology», New Yorker, 16 de noviembre de 2015, p. 52.

Sperber, D. (ed.), Metarepresentations: A Multidisciplinary Perspective, Oxford University Press, Oxford, 2000.

Sperber, D. y D. Wilson, Relevance: Communication and Cognition, Harvard University Press, Cambridge, 1986 (hay trad. esp.: La relevancia, Machado, Boadilla del Monte, 1994).

Sterelny, K., Thought in a Hostile World: The Evolution of Human Cognition, Blackwell, Malden, 2003.

—, The Evolved Apprentice, MIT press, Cambridge, 2012.

Strawson, G., Reseña de Freedom Evolves, de Daniel Dennett, New York Times Book Review, 2 de marzo de 2003.

Strawson, P. F., «Intention and Convention in Speech Acts», Philosophical Review, 73 (octubre de 1964), pp. 439-460.

Sullivan-Fedock, J., «Increasing the Effectiveness of Energy Wind Harvesting with CFD Simulation-Driven Evolutionary Computation», presentado en el congreso mundial de 2011 del Council on Tall Buildings and Urban Habitat, en Seúl, CTBUH, Seúl, 2011.

Swiss, J. I., «How Magic Works», 2007, http://www.egconf.com/videos/how-magic-works.

Szostak, J., «Systems Chemistry on Early Earth», Nature, 14 de mayo de 2009, pp. 171-172.

Tegla, E., A. Gergely, K. Kupan, A. Miklo y J. Topa, «Dogs' Gaze Following Is Tuned to Human Communicative Signals», Current Biology, 22 (2012), pp. 209-212.

Thomas, E. M., The Hidden Life of Dogs, Houghton Mifflin, Boston, 1993.

Thompson, D'Arcy W., On Growth and Form, Cambridge University Press, Cambridge, 1917 (hay trad. esp.: Sobre el crecimiento y la forma, Akal, Tres Cantos, 2011).

Tinbergen, N., The Study of Instinct, Clarendon Press, Oxford, 1951.

—, Social Behaviour in Animals, with Special Reference to Vertebrates, Methuen, Londres, 1959.

—, The Herring Gull's World; A Study of the Social Behaviour of Birds, Basic Books, Nueva York, 1961.

—, Animal Behavior, Time, Nueva York, 1965.

Tomasello, M., A Natural History of Human Thinking, Harvard University Press, Cambridge, 2014.

Tononi G., «Consciousness as Integrated Information: A Provisional Manifesto», Biological Bulletin, 215, 3 (2008), pp. 216-42.

Trivers, R., Social Evolution, Benjamin/Cummings, Menlo Park, 1985.

Turing, A. M., «On Computable Numbers, with an Application to the Entscheidungs Problem», Journal of Math, 58 (1936), pp. 345-363.

—, «Computing Machinery and Intelligence», Mind, 59 (1960), pp. 433-

460.

von Neumann, J. y O. Morgenstern, Theory of Games and Economic Behavior, Princeton University Press, Princeton, 1953.

von Uexküll, J., «A Stroll through the Worlds of Animals and Men: A Picture Book of Invisible Worlds», en Instinctive Behavior: The Development of a Modern Concept, ed. por C. H. Schiller, International Universities Press, Nueva York, 1934.

Voorhees, B., «Dennett and the Deep Blue Sea», J. Consc. Studies, 7 (2000), pp. 53-69.

Walsh, P. T., M. Hansell, W. D. Borello y S. D. Healy, «Individuality in Nest Building: Do Southern Masked Weaver (Ploceus velatus) Males Vary in Their Nest-building Behaviour?», Behavioural Processes, 88, 1 (2011), pp. 1-6.

Wegner, D. M., The Illusion of Conscious Will, MIT Press, Cambridge, 2002.

Westbury, C., y D. C. Dennett, «Mining the Past to Construct the Future: Memory and Belief as Forms of Knowledge», en Memory, Brain, and Belief, ed. por D. L. Schacter y E. Scarry, Harvard University Press, Cambridge, 2000, pp. 11-32.

Whiten, A., y R. W. Byrne, Machiavellian Intelligence II: Extensions and Evaluations, Cambridge University Press, Cambridge, 1997.

Wiener, N., Cybernetics: Or Control and Communication in the Animal and the Machine, Hermann/MIT Press, París/Cambridge, 1961 (hay trad. esp.: Cibernética, Tusquets, Barcelona, 1985).

Wills, T., S. Soraci, R. Chechile y H. Taylor, «'Aha' Effects in the Generation of Pictures», Memory & Cognition, 28 (2000), pp. 939-948.

Wilson, D. S., Darwin's Cathedral: Evolution, Religion, and the Nature of Society, University of Chicago Press, Chicago, 2002.

Wilson, R. A., http://www.rawilson.com/sitnow.html.

Wimsatt, W., y M. Beardsley, «The Intentional Fallacy», en The Verbal Icon: Studies in the Meaning of Poetry, University of Kentucky Press, Lexington, 1954.

Wrangham R., D. Cheney, R. Seyfarth y E. Sarmiento, «Shallow-water Habitats as Sources of Fallback Foods for Hominins», Am. J. Phys. Anthropol., 140, 4 (2009), pp. 630-642.

Wright, R., NonZero: The Logic of Human Destiny, Pantheon Books, Nueva York, 2000 (hay trad. esp.: Nadie pierde: la teoría de juegos y la lógica del destino humano, Tusquets, Barcelona, 2005).

Wyatt, R., y J. A. Johnson, The George Gershwin Reader, Oxford University Press, Nueva York, 2004.

Yu, Wei et al., «Application of Multi-Objective Genetic Algorithm to Optimize Energy Efficiency and Thermal Comfort in Building Design», Energy and Buildings, 88 (2015), pp. 135-143.

Zahavi, A., «Mate Selection—A Selection for a Handicap», Journal of Theoretical Biology, 59 (1975), pp. 205-214.

ÍNDICE DE ILUSTRACIONES

Figura 1.1. El conejo/pato

Figura 1.2. El cubo de Necker

Figura 3.1. Los círculos de piedras de Kessler y Werner

Figura 3.2. Algoritmo de ordenamiento de piedras de Stone y Kessler

Figura 3.3. Castillo de termitas australianas

Figura 3.4. Gaudí: La Sagrada Familia

Figura 4.1. Página de un manual de instrucciones para ascensoristas

Figura 5.1. Rastrillo para almejas

Figura 7.1. Espacio darwiniano

Figura 7.2. Espacio darwiniano con otras variables

Figura 7.3. Espacio darwiniano para el origen de la vida

Figura 7.4. Espacio darwiniano con religiones

Figura 7.5. Espacio darwiniano invertido, con fenómenos darwinianos en (0,0,0) y diseño inteligente en (1,1,1)

Figura 9.1. El gran árbol de la vida

Figura 9.2. Árbol glosogenético de las lenguas indoeuropeas

Figura 9.3. Adaptación al castellano del CAT de Selfridge

Figura 12.1. Patrones aleatorios de tetrominós evolucionando (Claidière et al.)

Figura 13.1. Espacio darwiniano

Figura 13.2. Espacio darwiniano de la evolución cultural, con fenómenos intermedios

Figura 14.1. Imagen sucesiva de colores complementarios

El título original de esta obra de Daniel C. Dennett es From Bacteria to Bach and Back. The Evolution of Minds.

El traductor quiere agradecer a Marià Pitarque su ayuda en la revisión de algunas secciones de esta obra.

Su primera edición en lengua inglesa fue publicada por W. W. Norton & Company.

Los derechos originales de esta obra pertenecen a:

© Daniel C. Dennett, 2017

Los derechos exclusivos de publicación en lengua castellana pertenecen a:

© Ediciones de Pasado y Presente, S.L., 2017

Pau Claris, 147, 4º, 1ª, 08010 Barcelona

ediciones@pasadopresente.com

www.pasadopresente.com

Esta edición de De las bacterias a Bach. La evolución de la mente ha sido compuesta en tipos Fournier y Gonzalo Pontón ha realizado la corrección de pruebas.

ISBN DIGITAL: 978-84-122888-6-5

ISBN: 978-84-947333-2-1

Depósito legal: B. 22.289-2017

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede realizarse con la autorización de sus titulares salvo en las excepciones que determina la ley. Si necesita fotocopiar o escanear fragmentos de esta obra, diríjase al Centro Español de Derechos Reprográficos (CEDRO) a través de la web www.conlicencia.com o mediante llamada telefónica al 91 702 19 70 o al 93 272 04 45

Índice

Cover Page 1
De las bacterias a Bach 2
PREFACIO 6
PRIMERA PARTE EL MUNDO, PATAS ARRIBA 10
1 INTRODUCCIÓN 11
2 ANTES DE LAS BACTERIAS Y ANTES DE BACH 31
3 SOBRE EL ORIGEN DE LAS RAZONES 41
4 DOS EXTRAÑAS INVERSIONES DEL RAZONAMIENTO 61
5 LA EVOLUCIÓN DE LA COMPRENSIÓN 84
SEGUNDA PARTE DE LA EVOLUCIÓN AL DISEÑO INTELIGENTE 109
6 ¿QUÉ ES LA INFORMACIÓN? 110
7 UN ENTREACTO DE ESPACIOS DARWINIANOS 141
8 CEREBROS FORMADOS POR CEREBROS 158
9 PALABRAS Y EVOLUCIÓN CULTURAL 183
10 EL PUNTO DE VISTA DE LOS MEMES 211
11 PROBLEMAS CON LOS MEMES: OBJECIONES Y RÉPLICAS 226
12 LOS ORÍGENES DEL LENGUAJE 253
13 LA EVOLUCIÓN DE LA EVOLUCIÓN CULTURAL 287
TERCERA PARTE NUESTRAS MENTES VUELTAS DEL REVÉS 337
14 LA CONCIENCIA COMO UNA EVOLUCIONADA ILUSIÓN DE USUARIO1 338
15 LA ERA DEL POSDISEÑO INTELIGENTE 375

ANEXO ALGUNOS ANTECEDENTES 418
NOTAS 437
BIBLIOGRAFÍA 459
ÍNDICE DE ILUSTRACIONES 492