

CONSCIENCIA

Cuando su padre le comunicó su alarma por haber olvidado hasta los hechos más impresionantes de su niñez, Aureliano le explicó su método, y José Arcadio Buendía lo puso en práctica en toda la casa y más tarde lo impuso a todo el pueblo. Con un hisopo entintado marcó cada cosa con su nombre: mesa, silla, reloj, puerta, pared, cama, cacerola. Fue al corral y marcó los animales y las plantas: vaca, chivo, puerco, gallina, yuca, malanga, guineo. Poco a poco, estudiando las infinitas posibilidades del olvido, se dio cuenta de que podía llegar un día en que se reconocieran las cosas por sus inscripciones, pero no se recordara su utilidad. Entonces fue más explícito. El letrero que colgó en la cerviz de la vaca era una muestra ejemplar de la forma en que los habitantes de Macondo estaban dispuestos a luchar contra el olvido: Esta es la vaca, hay que ordeñarla todas las mañanas para que produzca leche y a la leche hay que hervirla para mezclarla con el café y hacer café con leche. Así continuaron viviendo en una realidad escurridiza, momentáneamente capturada por las palabras, pero que había de fugarse sin remedio cuando olvidaran los valores de la letra escrita,

GABRIEL GARCÍA MÁRQUEZ (CIEN AÑOS DE SOLEDAD).

What is mind? No matter. What is matter? Never mind,

GEORGE BERKELEY.

Si no lo hubiera creído, no lo habría visto,

MARSHALL McLuhan.

Y seguimos subiendo por la escalera de la complejidad (Figura 91). Acabamos de llegar a un nuevo piso estable, generado por la emergencia de un fenómeno extraño y misterioso: la consciencia. Vamos a dar una breve in-

troducción a lo que se sabe acerca de ella, a algunos experimentos e hipótesis relevantes, pero no daremos una revisión histórica y quien esté interesado puede remitirse a libros de filosofía bastante agradables de leer como los de Daniel Dennett (1991, 1995, 1996 y 2003). La consciencia es algo tan subjetivo y misterioso que durante mucho tiempo la ciencia se negó a considerarlo como un tema de estudio. Pero si, como hemos visto, la inteligencia es algo rítmica, ¿no será que la consciencia es otro algoritmo? Enseguida haremos una propuesta en este sentido que, hasta donde yo sé, es completamente nueva y se diferencia radicalmente de las demás, pues muestra la consciencia como un fenómeno emergente bajo una interpretación computacional.

Comencemos pues, por los hechos históricos más relevantes.

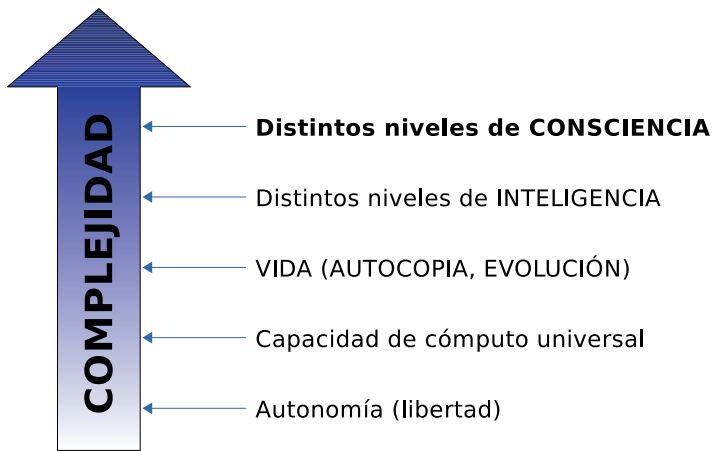


Figura 91. No se sabe cuántos bits se requieren para lograr consciencia, pero no son muchos, pues solo se precisa un bucle de realimentación

Se nos ha presentado siempre la consciencia como algo misterioso que está más allá del entendimiento humano. En cierto modo, así es. Se trata de una de tantas limitaciones *gödelianas* que tenemos. Recordemos que cuando Gödel logra codificar la aritmética de números enteros para que hable de sí misma, aparecen proposiciones que son verdaderas pero indemostrables. De manera análoga, cuando enfocamos nuestra inteligencia para que investigue sobre ella misma, lo logramos hacer, pero aparecen fenómenos nuevos que escapan a nuestra comprensión. La consciencia es real, está allí y la sentimos trabajar y, sin embargo, no entendemos qué es. De forma similar a las proposiciones de Gödel, para lograr comprenderla hay que dar un salto por fuera del sistema, que es lo que se presentará en este capítulo.

Ha habido escasos avances en la comprensión de la consciencia hasta hace relativamente poco. El dualismo cuerpo-mente de Descartes con sus

connotaciones religiosas ya está superado. Pero tuvo su importancia pues significó reconocer dos componentes, el *hardware* y el *software*, como diríamos hoy. El *hardware* se sabía que estaba formado por muchos módulos, como el corazón, el hígado o el páncreas. Pero se suponía que el *software* era monolítico, y solo muy recientemente se ha ido entendiendo que también consta de módulos. Actualmente, gracias a la tomografía cerebral en sus diversas variantes se han obtenido mapas del cerebro con las funciones específicas de cada parte así como correlaciones entre ellas y lo que decimos que estamos haciendo o sintiendo. Esto es lo que se llama la interpretación correlacional de la consciencia.

Incluso antes de disponer de estas herramientas electrónicas, las enfermedades y accidentes que dañaban una parte del cerebro permitían a los investigadores hacer deducciones basándose en las funcionalidades que se iban perdiendo.

Uno de los casos más emblemáticos fue el del obrero de ferrocarril Phineas Gage al que una varilla de hierro le atravesó la cabeza en un accidente en 1848. En el hospital le salvaron la vida, pero sus amigos decían que desde entonces ya no era el mismo. Se había convertido en una persona grosera y sin consideraciones para nadie. De allí y otros casos similares se dedujo que los dos lóbulos frontales, que fueron la región cerebral que sufrió todo el daño, eran los encargados de las funciones de alto nivel de la personalidad, como el autocontrol. Con ello se derrumba el mito de la unicidad de la consciencia y comienza a entenderse que el cerebro contiene módulos con funciones bien definidas y que el resultado de la operación conjunta de todos ellos es lo que produce inteligencia y consciencia. A partir de allí quedaron abiertas las puertas para que la ciencia investigase cuáles son esos módulos, cuál es su funcionalidad y cómo interaccionan entre sí.

Con estos hallazgos se afianza la psicología evolutiva cerrando un capítulo oscuro de un área de conocimiento, la psicología tradicional, que hasta ese entonces no llegaba a ser una ciencia. La psicología evolutiva sustenta que los módulos cerebrales aparecieron de forma evolutiva para que los genes pudieran solucionar problemas de supervivencia y reproducción en un entorno cada vez más complejo. Los genes fabrican cuerpos con cerebros que vienen ya preprogramados con algunos comportamientos simples de tipo “si se acerca un individuo muy grande y con la boca abierta, huye”, a los que llamamos instintos. Pero no pueden lidiar con comportamientos complejos en los que haya que analizar las sutilezas del entorno y tomar decisiones que dependen de múltiples condiciones. No hay genes que codifiquen “si se acerca una persona, y la persona es tu jefe, y está de mal humor, y hay rumores de despidos en tu empresa, entonces dile sí a todo”.

Uno de los pilares de la psicología evolutiva, como nos explica Steven Pinker en su libro *Como a mente funciona* es que cada módulo cerebral evolucionó para lidiar con estas sutilezas (si esa persona es o no tu jefe, cómo evaluar qué grado de humor tiene hoy o si los rumores que corren se pueden considerar fiables). Y, dado que los cambios culturales humanos son muy rápidos mientras que la evolución genética es lenta, es razonable suponer que los genes estén solucionando grandes problemas de hace miles de años, pero no sean capaces de lidiar correctamente con los de hoy en día. Por eso nos asustan instintivamente las culebras mientras que no le tenemos ningún miedo a los automóviles, a pesar de que las primeras ya prácticamente no causan víctimas entre la población humana mientras que los segundos son una de las principales causas de muerte en cualquier parte del mundo. Hay genes para reconocer y sentir miedo de serpientes y arañas, pero no ha habido tiempo de que se desarrollen genes que hagan lo mismo con los automóviles. Ni lo llegará a haber, porque cuando pasen los miles de años que requiere la evolución, para entonces las culturas humanas habrán inventado otros medios de locomoción irreconocibles en nuestro presente.

Lo anterior sucede también con muchas situaciones de la vida cotidiana. Los celos tuvieron razón de ser en los humanos (y siguen teniéndola en los animales) en una época donde nadie quería cargar con la alimentación y cuidado de unos hijos que potencialmente podrían ser de otro macho. Hoy día los análisis de ADN pueden resolver fácilmente esa cuestión sin necesidad de violencias, pero todavía sentimos celos porque esos genes siguen estando allí. Y se puede decir lo mismo de la mayoría de sentimientos negativos como la ira, la envidia o el impulso de amedrentar a otros. Los tribunales de justicia, la propia psicología evolutiva o la teoría de juegos, que son herramientas culturales humanas, vuelven inútiles estos sentimientos primitivos. Pero los genes no lo saben y eso crea conflictos mentales en las personas que no entienden que sus sentimientos están preprogramados para resolver problemas que ya no existen. Debería ser de cultura general tener unas bases mínimas en psicología evolutiva desde el colegio, para evitar mucha violencia y sufrimiento en la vida de adultos. Y debería ser obligatorio pasar exámenes al respecto si se aspira a representar un colectivo en algún tipo de organización económica, política o social, pues vemos en las noticias cómo muchos grandes dirigentes del mundo fallan estrepitosamente, comportándose como niños (o como simios) en discusiones que podrían resolverse de forma más civilizada.

Para ejemplificar voy a remitirme a un conocido presidente de un gran país, elegido hace pocos días. Resulta que este presidente toma decisiones de manera visceral, dictadas por su carácter, es decir, por sus genes. No

analiza primero pros y contras, no cuenta con la información adecuada de la situación actual y de las reglas del juego, y además su principal estrategia es repartir amenazas e insultos para todos, tanto a los críticos internos como a los externos e incluso a los amigos. Lo más probable es que no dure mucho así. En teoría de juegos se sabe que las amenazas tienen sentido si se posee la capacidad y la voluntad de cumplirlas. O, en caso contrario, si el oponente no sabe que te falta esa capacidad o voluntad. Pero las decisiones económicas proteccionistas de Trump (vaya, se me escapó) y sus correspondientes amenazas a diversos países no tienen ningún asidero, pues de llevarse a cabo el principal perdedor será su propio país. Nadie creará esas amenazas. Este no pasa de ser un ejemplo memorable, pero no es la excepción sino la culminación, pues muchos países tienen dirigentes con características similares. Es hora de decir ¡basta! Estamos ya en el siglo XXI, en la era de la información. Los conocimientos sobre la teoría de juegos y la psicología evolutiva están disponibles para todos los públicos, por lo que no deberíamos elegir a gente con tanta ignorancia, que incluso se vanaglorian de no leer ningún libro. Educar en estas áreas desde la niñez es muy importante, por el bien de la humanidad.

Volvamos al relato de casos espectaculares que revelan algún tipo de malfuncionamiento en los módulos cerebrales. La prosopagnosia es la incapacidad de reconocer rostros. Es una enfermedad relativamente común y sin mayores consecuencias, pero hay algo similar y mucho más inquietante que es lo que se llama el síndrome de Capgras, que se produce como consecuencia de la demencia senil o de golpes fuertes en la cabeza. El sujeto que lo padece cree que las personas de la familia han sido sustituidas por impostores, que son idénticos en apariencia. Lo que ocurre realmente es que la capacidad consciente de reconocimiento de rostros sigue intacta (no hay prosopagnosia), pero dejó de funcionar la capacidad de reconocimiento afectiva, que es inconsciente y que posiblemente esté mediada por los olores. En el libro *El hombre que confundió a su mujer con un sombrero* del neurólogo Oliver Sacks (2002) se comentan en detalle varios casos similares.

Otro caso que merece la pena comentar por lo sutil, es el del falso ciego (el término científico es anosognosia visual). El sujeto que lo padece alega que no puede ver. Según el relato de Dennett (1995, p. 371), lo que hizo sospechar a los médicos que ello podría ser un engaño fue que cuando al paciente se le ponía a caminar en una gran habitación vacía que solo tenía una silla, invariablemente se dirigía hacia ella, chocándose, y así demostrando (en su propio modo de entender la situación) que él estaba ciego. Obviamente, si se dirige siempre hacia la silla es porque sabe dónde está. La puede ver. El caso se resolvió cuando se dieron cuenta de que tenía dañada

la consciencia de la visión, no la visión en sí. De alguna manera podía ver a nivel inconsciente, pero no era consciente de ello. Y algo parecido nos pasa a todos, pues en cada ojo hay un punto bastante cerca del centro de la visión, que no tiene fotorreceptores debido a que es por allí desde donde sale el nervio óptico hacia el cerebro. Lo deberíamos ver como dos círculos negros en nuestro campo de visión, pero lo cierto es que nadie se da cuenta porque allí están suprimidos los circuitos neuronales que nos harían conscientes de ello. Como dice Dennett, “somos ciegos a la ceguera del punto ciego”.

Y el último ejemplo, quizás el más sorprendente de todos porque pone en duda quiénes somos, si somos libres y para qué sirve realmente la consciencia, también descrito por Dennett en el mismo libro (p. 177): a una persona se le han colocado electrodos en su cerebro, con su consentimiento, que miden la actividad motora de la mano derecha. Se le pide que pulse un botón de un proyector de diapositivas cuando en un reloj la aguja pase por las doce. Pero él no sabe que el botón del proyector realmente no hace nada, y que las diapositivas cambian cuando el electrodo detecta el comienzo de la actividad neuronal que va a contraer sus músculos de la mano para pulsar el botón. La persona se sorprende de que la diapositiva cambie justo antes de que ella tome la decisión de pulsar el botón. Le parece magia. El retraso medido desde que comienza la actividad neuronal hasta que la persona es consciente de que va a realizar la acción está entre 300 ms y 500 ms, es decir, ¡medio segundo! Pero entonces ¡la consciencia no es quien toma las decisiones!

Sin necesidad de hacer montajes tan invasivos, todos hemos experimentando este efecto cuando mantenemos una conversación informal en un ambiente agradable entre amigos. Uno habla y habla espontáneamente hasta que de repente dice algo inconveniente, o que se suponía era un secreto. Justo cuando tus palabras están saliendo de tu boca, te das cuenta de que has metido la pata. Esa es la consciencia, que no está a cargo de la conversación, sino solo supervisándola, y que da la orden de callarse demasiado tarde.

Estos experimentos y muchos otros ejemplos involucrando otras partes del cerebro están bien documentados en artículos científicos y en libros de divulgación como los de Damasio (2006), Dennett (1995) y Sacks (2002). Ello nos conduce a pensar que los módulos cerebrales evolucionaron para resolver problemas específicos que nuestros ancestros enfrentaron hace miles de años. Y que el cerebro no está diseñado para el arte o las matemáticas, de modo que si hemos logrado entenderlas es porque se trata de un efecto secundario producido por tener tantos módulos diversos con tanta capacidad de cómputo. Pero, además, ante ciertas situaciones no habituales para nuestros ancestros, los módulos pueden fallar estrepitosamente. Ejemplos divertidos de ello son las ilusiones ópticas (Díaz, 2015) y acústicas (Mikrojo,

2016a, 2016b, 2017a, 2017b), y también los errores en la forma de razonar de los humanos (Pinker, 1997, p. 357), como vimos en el capítulo anterior.

Entonces la mente está formada por muchos módulos, que pueden funcionar bien o mal. Uno de esos módulos es la consciencia y si reflexionamos nos damos cuenta que no siempre está presente. Cuando dormimos o estamos bajo anestesia, no somos conscientes. Y, cuando está presente, lo hace en diversos grados, por ejemplo, cuando vamos conduciendo un automóvil en dirección a un teatro a la vez que estamos sumergidos en una conversación con un acompañante, es posible que lleguemos a nuestra casa o trabajo, en vez de al teatro, para nuestra propia sorpresa. Nuestro cuerpo y nuestra mente inconsciente tomaron las riendas del automóvil y lo llevaron al “lugar de siempre”, sin que la consciencia interviniera para nada.

Por cierto que aunque un trozo de mi cerebro sea consciente y otro no, ambos forman parte integral de mí mismo. No hay que pensar en el inconsciente como algo o alguien externo, ajeno o incluso traidor a mí. Es también una parte de mí, pero de la que sé poco porque no puedo interrogarla. Una bonita conjetura acerca de la consciencia, alabada y criticada por igual, propuesta por Julián Jaynes y explicada en Dennett (1995), supone que ella es un fenómeno relativamente reciente entre los humanos: tiene entre 500 y 3000 años de antigüedad, lo cual es muy poco evolutivamente hablando, y podría decirse que estamos todavía en el proceso de su emergencia.

Y ello implica, aunque es políticamente incorrecto decirlo, que haya humanos más conscientes que otros, por motivos puramente genéticos, de la misma manera que los hay más inteligentes que otros por los mismos motivos. La hipótesis de Jaynes, conocida como “de la ruptura de la mente bicameral”, señala que los dos cerebros que tenemos hace miles de años estaban bastante más desunidos que ahora, y solo había un flujo efectivo de información del lado derecho al izquierdo. Recordemos que los módulos del habla se encuentran principalmente en el cerebro izquierdo de las personas diestras. Además, cada lado del cerebro recibe información sensorial y controla los músculos del lado contrario del cuerpo.

Uno de los indicios a favor de esta hipótesis es que la palabra “consciencia” con su significado moderno solo existe desde hace unos 500 años. Por supuesto, uno puede aprender a usar unos palillos chinos en China sin saber cuál es su nombre. Pero se hace muy difícil de sostener lo mismo para la consciencia, puesto que ser consciente implica la capacidad de reflexión sobre uno mismo y sobre su consciencia. Tener una palabra para ello es ineludible. El neurólogo Mariano Sigman (2016) ha realizado un análisis automático del uso de la palabra consciencia y de otras palabras relacionadas, en textos desde la antigua Grecia hasta hoy, corroborando esta hipótesis.



Fuente: Rebecca Goldstein (2005). Disponible en: <https://goo.gl/SERxVN>

Personaje 8

Steven Pinker (1954-)

Steven Pinker es un lingüista y psicólogo evolutivo canadiense. De hecho, sus dotes como escritor divulgador de ciencia han logrado llevar al público la idea de que la evolución es la única vía para darle sustento científico a la desprestigiada psicología tradicional. Según Pinker, la mente tiene módulos funcionales que han evolucionado para responder a las necesidades de los humanos y, entre ellos, destaca el módulo del lenguaje, tal y como decía Chomsky. Con ello se opone frontalmente a los psicólogos tradicionales que consideran al cerebro de un niño como una tabla rasa, en la que se puede construir cualquier cosa por medio de un aprendizaje adecuado. Por el contrario, Pinker piensa que el cerebro de un niño al nacer viene dotado de una serie de capacidades innatas que pueden ser aprovechadas o no, por medio de aprendizaje, pero que, debido a ello hay cosas que puede aprender (como cualquier lenguaje humano, cosa que el cerebro de los demás primates no puede), hay cosas que no puede aprender, y hay cosas que puede aprender con dificultad (como las matemáticas avanzadas), en función de los módulos innatos que posea. Otra conclusión interesante es que los 10 000 años que llevamos desde que existe la agricultura y los 200 desde que existe la industria no son suficientes para cambiar los genes, de modo que nuestro carácter y nuestra forma de pensar es adecuada para los problemas que teníamos en el Pleistoceno. Otros científicos que trabajan o apoyan la psicología evolutiva son David Buss, Leda Cosmides, Martin Daly, Margo Wilson, Daniel Dennett, Richard Dawkins y el divulgador científico Roberth Wright, aunque también hay críticos que piensan que debe reformularse con más rigor, como Buller (2005).

También son muy interesantes sus trabajos en lingüística, donde consigue armar un árbol evolutivo de todas las lenguas humanas, que coincide muy bien con las migraciones, y hacer arqueología de palabras. Una de las ideas principales es que las sentencias de cualquier lenguaje (incluyendo los de computador) son árboles, pero hay que serializarlos para convertirlos en habla. El proceso de conversión depende de cada lenguaje y allí identifica unos 20 parámetros que controlan cómo hacerlo (si el verbo va al final o separando sujeto y complemento; o si el adjetivo va antes o después del nombre). Los niños, al aprender la lengua de sus padres, además de identificar el vocabulario lo que hacen es ajustar esos 20 parámetros.

Otro de los indicios son los escritos de la Grecia clásica. En la *Iliada* y la *Odisea*, Homero dice que sus personajes escuchaban las voces de las musas y de los dioses. Si vivieran hoy, serían diagnosticados de esquizofrenia. Pero en aquella época podría significar que el hemisferio derecho estaba comunicándose con el izquierdo, que es el que tiene capacidad de hablar al mundo exterior. El izquierdo es el “*front-end*”, como se dice en computación,

y explica a las demás personas que está oyendo voces, aun cuando no sabe de dónde, porque no tiene un bucle hacia atrás que le permita dialogar y conocer nada acerca de su lado derecho.



Fuente: CC BY 2.0, David Orban (2006). Disponible en: <https://goo.gl/NDKNtR>

Personaje 9

Daniel Dennett (1942-)

Daniel Clement Dennett es un filósofo extraordinario y un gran divulgador. La filosofía, que al principio fue el motor de las ciencias, con el tiempo se ha convertido en un enredo de palabras sin sentido. Hay unos pocos filósofos que se apartan de ese esquema y entre ellos sobresale Dennett, que siempre trabaja de la mano con científicos, neurólogos e ingenieros. Es decir, hace filosofía pero partiendo de datos experimentales. Por el contrario, se dice que los filósofos tradicionales tienen los pies firmemente asentados sobre el aire.

Entre los libros de Dennett destaca *La peligrosa idea de Darwin*, donde explica la evolución y pone múltiples ejemplos, no solo en el ámbito biológico. Todavía hay un cierto debate entre legos sobre si la evolución es verdad o no. A ellos les recomiendo que aprendan a programar y escriban un algoritmo evolutivo, con lo que se les despejarán las dudas para siempre de lo bien que funciona. Otros libros interesantes son *Tipos de mentes*, *La consciencia explicada* y *La libertad evoluciona* donde elimina muchos mitos sobre estos temas, dejándonos la idea de que el yo, los qualia y la libertad, son solo una ilusión. Un libro anterior, *The mind's I*, escrito con Douglas Hofstadter, explora muchos tópicos sobre la consciencia de una manera inusual: primero incluyen un texto de un reconocido autor como Borges, Turing, Dawkins, Smullyan, Searle y otros, y luego lo analizan o critican, argumentando para alcanzar su objetivo final, que es demostrar que la mente se puede mecanizar, es decir, se puede lograr la inteligencia artificial en sentido fuerte. Recuerdo especialmente los cuentos de ciencia ficción de Stanislaw Lem, terriblemente divertidos y anticipadores.

Daniel Dennett amplía esta hipótesis respecto a la capacidad del habla humana: un protohomínido grita para atraer la atención de otros de su misma especie, quizás porque encontró comida y necesita ayuda, quizás para avisar de algún peligro. En alguna ocasión, está solo y no hay nadie que responda a sus gritos. Pero de todos modos son oídos por su oreja izquierda, conectada a su cerebro derecho que puede reaccionar, tratando de ayudarlo. Este puente de comunicación externo izquierdo → derecho, con el tiempo podría irse interiorizando, ya que tendría obvias ventajas: conseguir que la comunicación fuera privada. De modo que podría haber presión evolutiva empujando en esa dirección (aun hoy día todos conocemos de personas que

necesitan decirse las cosas en voz alta para lograr ordenar sus pensamientos). Finalmente, con la comunicación bidireccional interna terminada, aparece un nuevo bucle que, cómo no, va a producir la emergencia de nuevas e interesantes propiedades: la consciencia.

Porque la consciencia es muy rica y tiene muchas facetas y grados de profundidad:

- **Consciencia del entorno.** Puedo usar los sentidos para percibir lo que me rodea, y detectar situaciones beneficiosas y perjudiciales, para acercarme a las primeras y alejarme de las segundas. Para ello hay que tener una memoria donde almacenar y clasificar situaciones anteriores similares. Además está la integración sensorial: cuando oímos una campana y vemos una campana, sabemos que se trata del mismo objeto. Y relacionado con ello está la capacidad de atención, es decir, poder centrar la consciencia en un único objeto en cada momento, ignorando todo lo demás. La mayoría de animales tienen este tipo de consciencia.
- **El concepto del yo.** Hay algo misterioso en ello, pero la idea principal es que puedo identificarme a mí mismo, separándome del entorno. Aparentemente, el *yo* es quien gobierna las acciones del cuerpo, ejerciendo su libertad (enseguida veremos que eso no es realmente así). Muchos filósofos distinguen varios matices en este concepto: el *yo*, el *ego*, el *mí mismo*⁸⁴... Nosotros los englobaremos en una sola cosa. Los animales mamíferos suelen tener una idea del *yo*. Se ha comprobado que si se le da un espejo a un orangután (que no ha visto tal artefacto en su vida), en poco tiempo lo usa para observarse a sí mismo, hacer muecas, ayudarse de ello para rascarse la espalda y cosas así. Ello significa que se reconoce en su reflejo, por lo que tiene un concepto de él mismo. Otros animales no son así, y se asustan al verse reflejados, pues creen que se trata de un rival.
- **Autoconsciencia.** Soy capaz de pensar sobre mí mismo y hacer introspección y análisis de lo que pienso y siento, de lo que hice o pude hacer o haré.
- **Los *qualia***⁸⁵. tengo una representación interna de mis sensaciones que es subjetiva, que no puede compararse con las de otros seres similares. Ello es algo bastante misterioso. Siempre se pone como ejemplo “la sensación que tenemos del color rojo”. No la podemos explicar. Es completamente nuestra. Y no sabemos, ni podremos saber de ninguna

84 Y de donde deriva una palabra tan interesante como “ensimismarse”.

85 *Qualia* es una palabra que viene del latín, está en plural, y su singular es *quale*.

manera, si cuando dos personas miran el mismo objeto rojo experimentan la misma sensación en su mente. Lo mismo puede decirse de la sensación de hambre, la sensación de llegar tarde, la sensación de estar cansado o feliz y tantas otras.

- **La subjetividad de la consciencia.** Está muy ligada a lo dicho sobre los *qualia*, pero es más general. Es la sensación misteriosa de estar experimentando un *yo* sumergido en un mundo de sensaciones que, aunque están relacionadas con parámetros que pueden medirse (si estoy en la calle a -20 grados experimento frío), la relación de mí con el mundo es completamente personal, indescriptible e incomparable con la de otras personas. Podemos sentir. No es que estemos viendo una película, es que estamos dentro de la película experimentando un flujo continuo de sensaciones. Es lo que el filósofo David Chalmers (1995) llama el problema principal de la consciencia⁸⁶.

La primera de estas características es la más abordable desde un punto de vista computacional y ya hay muchos trabajos que apuntan a resolverla, principalmente en el área de visión computacional. Los problemas que involucra son meramente técnicos. Pero las otras tres características tienen un halo de misterio que impide un acercamiento ingenieril. Sin embargo, aquí presentaré unas ideas que podrían cambiar esa situación.

Durante mucho tiempo se pensó que solo los humanos tenían consciencia, pero en los últimos años se ha revisado esta idea. Los animales tienen consciencia en mayor o menor grado. Recomiendo volver a ver la excelente presentación de Frans de Waal en TED (2011) donde explica que los fundamentos de la moralidad humana ya se encuentran en muchos primates y que la causa es evolutiva. Los chimpancés pueden cooperar y exigir cooperación a otros, así como los elefantes, que vuelven periódicamente a los sitios donde hubiera muerto algún miembro querido de su familia. Los delfines jóvenes se agrupan para molestar y golpear a algún otro; hacen *bullying*, como los adolescentes humanos. Y quienes tienen algún animal doméstico saben perfectamente que sienten, sufren, se alegran igual que su dueño. Los perros en algunos aspectos son muy poco inteligentes, pero en otros pueden demostrar comprensión y voluntad para realizar acciones que no son mecánicas sino que identificamos en ellas que “hay alguien allí dentro, decidiendo”. Ya se han hecho escáneres cerebrales de perros donde se descubre que, cuando están alegres, están pensando en sus amos. Algo similar ocurre con gatos y algunos pájaros como los cuervos, de lo que te das cuenta si eres buen ob-

⁸⁶ *The hard problem of consciousness.*

servador. La consciencia en sus múltiples facetas es una cuestión de grado, y no de todo o nada, un tema recurrente a lo largo de este libro.

Ha habido varios intentos para tratar de explicar de dónde surge el *yo*. Recuerdo muy bien el primero que llegó a mis oídos: se había descubierto una oscilación en el cerebro, en forma de un barrido de disparos neuronales en el córtex, de una frecuencia en torno a los 40 Hz. El famoso biólogo co-descubridor de la estructura del ADN, Francis Crick, proponía en 1990 que ese barrido de 40 Hz era el que producía la integración sensorial y daba lugar a la experiencia unificada subjetiva de la consciencia.

Eso da que pensar. ¿40 Hz es un número mágico y si fuera 258 Hz ya no funcionaría? ¿Si diseño un circuito electrónico que oscile a esa frecuencia tendrá consciencia? ¿O será que hay una reacción química clave en el cerebro que es la que produce la consciencia? La respuesta es obviamente negativa a todo ello. Tiene mucha razón Chalmers en sus preguntas (no tanto en sus respuestas) acerca de que no existe ningún fenómeno físico o químico que tenga la potencia causal para argumentar que genera consciencia. Estamos hablando de niveles distintos. No hay forma de argumentar que un barrido eléctrico sincronizado o una reacción química puedan producir algo tan extraño como la consciencia. Y eso me llevó a pensar: ¿qué tipo de razonamiento sería el adecuado, convincente, y tendrá la suficiente capacidad explicativa para lograrlo? Tiene que ser algo bastante distinto, no físico ni químico, quizás matemático, quizás sistémico como los bucles de realimentación, quizás computacional.

Por cierto que los 40 Hz lo que más bien indican es que el cerebro es digital a un cierto nivel. Obviamente, en el bajo nivel los potenciales de acción de cada neurona cambian de manera continua, como cualquier voltaje, pero la integración de la información sensorial y su percepción a alto nivel es discreta. Cada 25 milisegundos (la inversa de 40 Hz) el cerebro muestrea sus sensores y toma decisiones. Por ello, los sonidos cuya frecuencia es mayor a 20 Hz se perciben correctamente, mientras que en los de frecuencia menor se percibe cada onda como un evento individual, como un golpeteo. Y también por ello las películas de cine deben ir a 24 fotogramas por segundo (o más rápido) pues para cadencias más lentas comienzan a notarse las fotos de forma aislada y se pierde la sensación de continuidad. Aunque estos números no coinciden exactamente, si lo hacen de forma aproximada. Hay que tener en cuenta que la frecuencia de 40 Hz no es fija y puede llegar hasta 70 Hz, variando según nuestro estado de atención y también entre una persona y otra. Además, hay varios circuitos cerebrales involucrados y cada uno puede tener una frecuencia distinta. Laura Spinney (2015) nos ofrece una explicación más detallada de todo ello.

UNA PROPUESTA COMPUTACIONAL PARA ENTENDER LA CONSCIENCIA

En este estado de cosas podemos usar los conceptos de la vida artificial como vía para proponer una respuesta. Concretamente podemos imaginar qué le pasaría a un futuro robot que hubiera sido fabricado como resultado de un proceso evolutivo, en el sentido de que tiene motivaciones cableadas en *hardware* para sobrevivir, buscar situaciones beneficiosas y huir de los peligros. Vamos a imaginar que ya tiene inteligencia suficiente para identificar objetos en su campo visual, memorizar situaciones y hacer predicciones sobre el futuro. Todo esto son cosas técnicas que hoy día ya se están logrando y no hay allí ninguna magia ni misterio. Por ejemplo Google ya ha logrado un *software* de *deep learning* capaz de identificar gatos en videos, sin ayuda externa ni entrenamiento supervisado (Quoc *et ál.*, 2012).

La consciencia se ha entendido desde hace tiempo que debe basarse en algún tipo de bucle, como por ejemplo Edelman (1998) con sus mapas neuronales reentrantes y Hofstadter (2007) con sus bucles autorreferenciales. Ya adelanto que esta nueva propuesta se basa en un bucle con capacidad de predicción. Vamos a argumentar que la consciencia puede emerger a partir de módulos que intentan predecir. Recordemos del capítulo anterior que la inteligencia es principalmente la capacidad de predecir, y suponemos que el robot ya tiene ese nivel de inteligencia. Básicamente, el *yo* es un punto fijo de un conjunto de predictores realimentados sobre sí mismos. Y las sensaciones subjetivas son fallos de codificación. Estas ideas ya las publiqué en (García, 2017 y 2018), pero quisiera aquí mostrarlas más despacio, con más detalles y argumentaciones, cosa que no permiten los artículos.

Para empezar debemos saber que la zona más externa del cerebro humano, llamada neocórtex (de 2 mm de grosor), está conformada por medio millón de columnas corticales de 0,5 mm de diámetro, a su vez organizadas en 6 capas de neuronas. La arquitectura es tan repetitiva que ha llamado la atención de los científicos que trabajan con el cerebro (DeFelipe, 2012). De alguna forma ello está diciendo que el mismo algoritmo que se emplea para percibir sonidos, también se emplea para la visión, para razonamiento espacial, para la lógica y para más cosas. Y ya se sabe cuál es ese algoritmo tan general: cada columna cortical constituye un módulo dedicado a predecir las señales de entrada que le llegan a sus neuronas. Usando una implementación computacional de esta arquitectura, Jeff Hawkins (2004) ha creado un cerebro artificial que está disponible gratuitamente en la web. Y ha cosechado ya algunos éxitos interesantes haciendo, entre otras cosas, que aprenda canciones a base de repetírselas, como nos ocurre a los humanos. En EVALAB hemos usado ese cerebro artificial para que aprenda a cumplir

unas órdenes básicas en un entorno espacial (Narváez, 2016; Narváez, García y Gutiérrez, 2017).

Ahora debemos recordar qué es un punto fijo en matemáticas: si tenemos una función cuya salida se realimenta sobre la entrada, es decir:

$$\begin{aligned} y &= f(x) \\ x &= y \end{aligned} \qquad \text{Ec. 36}$$

Se dice que x_F es un punto fijo de esa función si

$$x_F = f(x_F) \qquad \text{Ec. 37}$$

Para garantizar que exista un punto fijo, la función debe ser continua y contractiva (Wikipedia, 2016b). Y el conjunto de entrada y de salida debe ser el mismo y estar acotado.

Los puntos fijos (Wikipedia, 2016a) son invariantes en un sistema y para entenderlos bien propongo que resolvamos primero el problema 21 de ingenio.

Problema 21: Punto fijo de Palmira a Tuluá

Luisa sale de Palmira a las 8:00 del lunes en su automóvil, y llega a Tuluá a las 13:00. El martes hace el recorrido inverso por la misma carretera, saliendo de Tuluá a las 8:00 y llegando a Palmira a las 13:00. La pregunta es: ¿existe algún punto de la carretera por el que Luisa pasó a la misma hora el lunes y el martes?

Ten en cuenta que no sabemos a qué velocidad circulaba Luisa y que ni siquiera era constante. Seguramente había semáforos, a veces en rojo, a veces en verde. Es posible incluso que en algún momento llevase una velocidad negativa (por ejemplo, si tuvo que dar reversa para recoger algo que olvidó). Lo único que sabemos es que a la ida y al regreso fue por la misma carretera y coincidieron las horas de salida (8:00) y las de llegada al destino (13:00).

La mayoría de la gente piensa que es muy improbable que exista ese punto, pero...

Causa bastante sorpresa este tipo de problemas la primera vez que te enfrentas a ellos. ¿Cómo es posible demostrar que exista algo si no se sabe dónde está? Incluso cuando ya conoces la solución y la has entendido bien, el siguiente problema que te pongan te puede volver a desconcertar (por favor, intenta resolver el problema 22).

Problema 22: Arrugando papeles

Supón que tienes una caja con papel A4 abierta por arriba. Coges el primer papel de todos, lo arrugas de cualquier manera (sin romperlo) y lanzas la pelota de nuevo sobre la caja. Y aquí la pregunta es: ¿existe algún punto del papel arrugado que esté en la misma vertical que estaba antes de arrugarlo?

De nuevo, dado que se puede arrugar de miles de maneras, y se puede lanzar a cualquier sitio encima de la caja, la respuesta es que parece muy improbable que algún punto vaya a permanecer en la misma vertical, pero...

¡Otra vez! ¿Cómo es posible? Parece magia que se pueda llegar a demostrar algo cuando se tiene tan poca información de las condiciones, en este caso, de cómo se arrugó el papel.

Hay una multitud de problemas similares y todos nos causan estupor y asombro. Cada uno plantea algo raro, único y aparentemente desconectado de los demás (Problema 23). Otro interesante y más complicado: busca por Internet el “teorema del no pelo” que demuestra que es imposible peinar por completo una pelota de tenis que tuviera pelo por todos los lados, y que se aplica no a la peluquería sino a algo tan distinto como los agujeros negros en la relatividad general.

Problema 23: La vela

El siguiente problema es para mostrar que los puntos fijos tienen aplicaciones prácticas. Tienes una vela que tarda en arder exactamente una hora. Pero la velocidad con que lo hace no es uniforme pues hay zonas más gruesas que otras. ¿Cómo puedes hacer para cronometrar exactamente media hora?

Volviendo al tema principal, como bien dicen Daniel Dennett y Jeff Hawkins, si tuviéramos un robot con un cerebro sin consciencia pero con percepción sensorial del mundo, las manos, los pies y todos los miembros y sensaciones de su cuerpo formarían parte de su entorno externo. El robot intentaría predecir todo lo que ve y siente, pues si lo logra anticipar podrá salvarse de peligros (hemos supuesto que el robot es el resultado de algún proceso evolutivo, por lo que tendrá cableado en *hardware* todo tipo de algoritmos de supervivencia). También hemos supuesto que el robot ya tiene capacidad para segmentar lo que percibe clasificándolo como objetos de distintos tipos (algo que hacen ya los algoritmos de Google que identifican gatos en videos).

Entonces, el robot se moverá por su mundo, percibiendo objetos y tratando de predecir con mayor o menor acierto lo que van a hacer, cuando de repente ¡algo extraordinario ocurre!: hay un objeto que siempre puede predecir bien al 100%. Si piensa que va a levantar la mano, efectivamente levanta la mano. Si piensa que va a dar un salto, o a hacer un gesto con la cara, efectivamente ello ocurre. Intenta desafiarlo tratando de hacer predicciones sobre movimientos extraños y repentinos⁸⁷ pero siempre acierta. Acaba de descubrir el yo, en su nivel más básico.

⁸⁷ Lo cual seguro que ya habrás experimentado cuando eras un adolescente y estabas delante de un espejo.

Recuadro 8: El "yo" de la consciencia

El yo es un punto fijo que aparece cuando un conjunto de funciones de predicción actúan sobre sí mismas. Y, como tal abstracción matemática, no tiene una existencia material. Es solo una ilusión.

Y causa asombro, claro, como todos los puntos fijos. Y es difícil de entender, y es misterioso... Pero, en fin, es solo eso: un punto fijo.

También es importante entender que mis acciones las decide todo el sistema nervioso incluyendo mi cerebro, mi cerebelo, mi tallo encefálico y otros centros nerviosos. Habitualmente achacamos al yo la capacidad de decidir y realizar acciones, ser el origen de la voluntad, ser el conductor que guía al resto del cuerpo, pero ello es incorrecto. Si nos confundimos con ello posiblemente sea debido a que hay una presión, en buena medida evolutiva, por encontrar relaciones de causa-efecto ya que eso facilita la capacidad de predecir el futuro. Y si el cuerpo se mueve mientras ese punto fijo que es el yo refleja esos movimientos con total exactitud, es razonable cometer el error de pensar en el yo como la causa. Pero los experimentos que menciona Dennett, que ya hemos explicado al principio del capítulo, demuestran que eso no es así. Las decisiones se toman de forma inconsciente por todo el cerebro y el resto del sistema nervioso. La consciencia las registra después de que se han tomado. Parece ser que la consciencia puede tener también una cierta capacidad de veto, es decir, de abortar decisiones que están comenzando a ejecutarse. Pero no decide las acciones.

El hecho de que estos módulos puedan predecir al 100% lo que va a hacer el resto del cuerpo ocurre porque el cerebro maneja información privilegiada: en algún sitio del cerebro y el cerebelo está la información motora, que va a enviarse hacia los músculos. Y conjeturo que los módulos de predicción pueden "hacer trampa" (línea roja de la figura 92), obteniendo de allí la información.

También quiero advertir que, a pesar de ello, no es posible hacer una demostración de la existencia de este punto fijo, pues la función de predicción no es continua (el espacio de objetos sobre la que opera es discreto y muy fraccionado). Suena razonable que un cerebro de robot cuya misión es predecir todo, tarde o temprano se encuentre a sí mismo dentro de ese todo. Pero no hay garantías, pues el cerebro podría tener defectos o peculiaridades que lo impidan. Es decir, pueden existir cerebros sin punto fijo y también cerebros con más de un punto fijo. Ello seguramente está siendo identificado ya como algún tipo de enfermedad, incapacidad o peculiaridad mental.

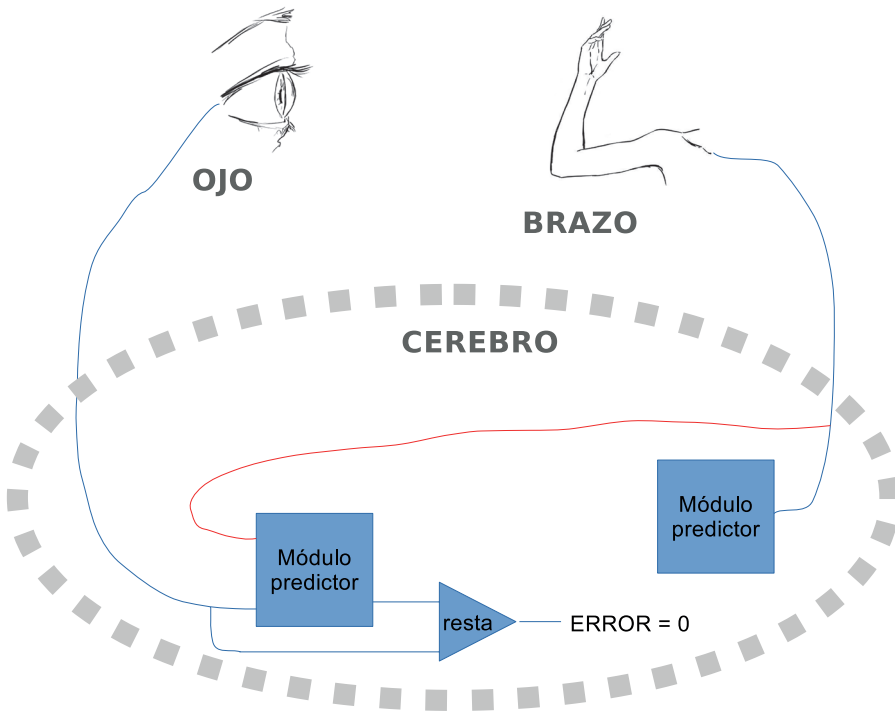


Figura 92. Bosquejo del bucle de predicción con error 0

Es razonable esperar que la presión evolutiva empuje a tener un único *yo* y hay varias razones para ello. Primero, este modelo de mí mismo produce un *yo*, que no solo es el centro de una experiencia narrativa, como dicen Dennett y Bruce Hood (2012), sino que también es a quien debo de salvar en caso de peligro. Ello tiene también obvias ventajas evolutivas. Y segundo, si gracias a la consciencia tengo un modelo de mí mismo (modelo que usa mi mente *popperiana*, como explicábamos en el capítulo anterior), entonces ese modelo también puedo aplicarlo a los otros individuos de mi misma especie, con lo cual puedo predecirlos para saber en quién confiar, con quién puedo cooperar y de quién debo cuidarme. Esta es la hipótesis del psicólogo Nicholas Humphrey (citado por Lewin, 1992, p. 158) acerca de la ventaja de ser consciente en un entorno social. Es lo que se llama tener una teoría de la mente de los otros.

Como novedad quisiera extender la hipótesis de Humphrey en el sentido de que hay una causalidad circular (un bucle de realimentación positiva): al jugar juegos sociales intento predecir a los otros y entonces emerge un objeto especial con el que siempre acierto, el *yo*. Con ello también emerge la autoconsciencia.

Hay un divertido experimento ideado en 1998 por Matthew Botvinick y Jonathan Cohen, de la Universidad Carnegie Mellon y replicado muchas veces (Bernardo, 2013) que corrobora esta propuesta. En este experimento se le pide a una persona que extienda los dos brazos encima de la mesa. El brazo izquierdo queda a la vista mientras que hay una pequeña pantalla que impide a la persona ver su brazo derecho. Al otro lado de la pantalla, y a la vista de la persona, colocan un brazo de plástico. Es obvio que es de plástico y la persona lo sabe. El experimentador se sienta enfrente de la persona, al otro lado de la mesa. Con ayuda de un pincel acaricia la mano de plástico a la vez que, al otro lado de la pantalla y sin que lo vea la persona, con otro pincel acaricia la mano derecha real. La persona siente el cosquilleo en su mano derecha a la vez que ve la mano derecha de plástico ser acariciada con el pincel y de allí llega a algo asombroso: identifica la mano de plástico como si fuera parte del *yo*.

Hay que recalcar esto: no se trata de que la persona sienta un fenómeno curioso de coincidencia sensorial. Es que, al ver la mano de plástico siendo acariciada puede predecir exactamente que va a sentir un cosquilleo; y al sentir el cosquilleo puede predecir con total exactitud que la mano de plástico está siendo acariciada. Recordemos que según esta propuesta, todo lo predecible con 100% de exactitud pasa a denominarse el *yo*, y efectivamente eso es lo que ocurre: la mano de plástico pasa a ser parte del *yo*.

El experimento termina de una manera divertida: el experimentador saca repentinamente un martillo y golpea la mano de plástico. A lo que la persona reacciona visiblemente con susto, gritando o intentando retirar la mano de allí (obviamente solo logra retirar la mano real). Ello corrobora de manera muy efectiva que la persona realmente ha identificado la mano de plástico como una parte de su *yo*, que quiere defender de la agresión.

Hay más experimentos donde al sujeto se le induce a creer que un objeto externo forma parte del *yo* como, por ejemplo, usar un lapicero para tocar objetos de distintas texturas, con los ojos cerrados. Por un lado, es fácil adivinar de qué objetos se trata, a pesar de no estar tocándolos directamente. Y, por otro lado, nos identificamos con la punta del lapicero: nosotros estamos allí, en la punta del lapicero. Lo mismo ocurre cuando nos volvemos expertos en el manejo de una herramienta, y la más cotidiana es el automóvil. Podemos sentir el tipo de terreno que tocan las ruedas (un bache, una piedra...) como si estuviéramos palpándolo con las manos. Podemos también sentir la extensión del automóvil, sabiendo con exactitud si cabe o no cabe entre dos obstáculos. En resumen, si podemos predecir al 100% cómo se va a comportar el automóvil, entonces lo incorporamos al *yo*.

La única crítica seria que he recibido hasta ahora es que los puntos fijos son atractores (o repulsores) de modo que el *yo* debería quedarse atrapado observando al *yo* para siempre. Es muy fácil resolver esta objeción cuando uno se da cuenta que el automóvil del problema 21 tampoco se queda atrapado para siempre en el punto fijo. Hay que distinguir entre el concepto matemático de punto fijo, que es el que estamos tratando aquí, y el concepto en el ámbito de los sistemas dinámicos que, efectivamente, genera atractores, tal y como vimos en el capítulo “Caos” del libro anterior.

Sin embargo, también hay que considerar el cerebro como un sistema dinámico, y sabiendo que un objetivo importante de cualquier proceso de predicción es disminuir su error, entonces el *yo* sí podría convertirse en un atractor en el que, una vez que piensas, no sales jamás.

Esto se puede rebatir. Por un lado, el cerebro no es solo el *yo*. Dado que tiene muchos módulos que trabajan en paralelo, puedo realizar varias tareas a la vez, como conducir un automóvil y escuchar música. Pero definitivamente la consciencia se dirige solo a una cosa a la vez. Hay un circuito de atención (que se puede modelar como en la figura 93), que se centra en uno de los muchos estímulos o pensamientos que lleguen a ese nivel. Y debe haber una funcionalidad que decide qué atender para la que ya hay algunos modelos como el de van den Boogaard y Turpijn (2017). Seguro que tienen prioridad las situaciones de supervivencia. Y si no hay nada de esta índole que atender se podrá conmutar de forma más o menos aleatoria entre todas las posibilidades. En algún momento caeremos en los módulos cuyo error de predicción es cero, pero el conmutador aleatorio nos sacará de allí antes o después. En principio eso no debería generar ningún bucle infinito, salvo que uno se aisle sensorialmente y se concentre mucho en pensar sobre el *yo* o la consciencia. O en alguna situación donde pensar en el *yo* sea vital para algo nuestro (como, por ejemplo, hacer experimentos sobre la percepción del *yo*). Ese tipo de actividades no son aconsejables pues pueden dañarnos la mente, dejándola atrapada, efectivamente, en un bucle de realimentación positiva potencialmente destructivo, en el que cuanto más pensamos en el *yo*, más queremos pensar en el *yo*.

De hecho, los experimentos de privación sensorial (donde se deja flotando al sujeto en un baño a la misma temperatura corporal, sin ningún ruido, luz o estímulos externos), si se prolongan demasiado tiempo pueden producir esos daños en el cerebro, porque el circuito de atención no tiene en qué otra cosa sintonizarse sino en uno mismo. Podemos ver un experimento interesante en Vsauce (2017), que es un excelente canal de divulgación científica.

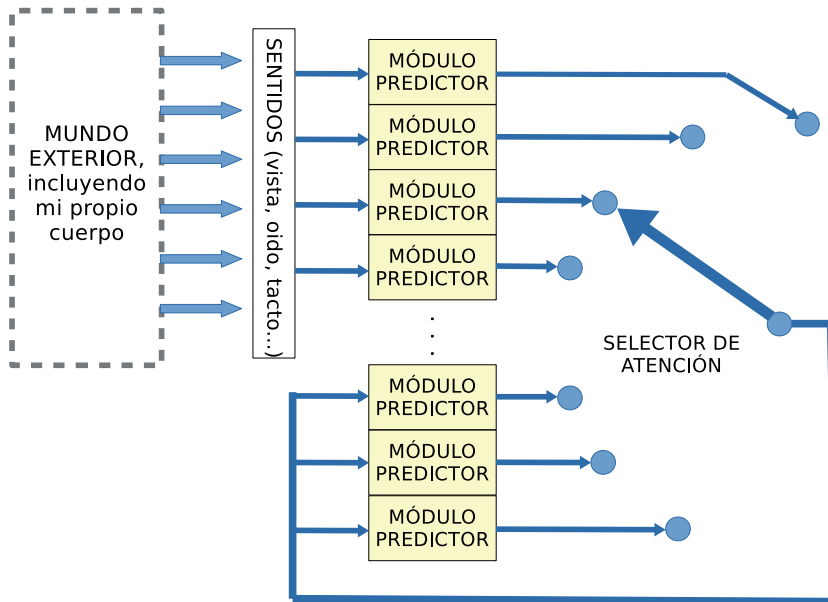


Figura 93. El selector de atención de la consciencia

La idea de que el circuito de atención puede saltar a otras cosas distintas al punto fijo, o sea, que el punto fijo puede desconectarse, se puede intuir a través de una analogía con el proceso de arrugar un papel, que veíamos en el problema 22. Como permitimos que el papel arrugado caiga en cualquier sitio, y hacemos lanzamientos cada cierto tiempo, la mayoría de las veces caerá fuera de la bandeja, y entonces no habrá ningún punto fijo. O sea, la mayoría de las veces la atención del cerebro estará fijada en cosas externas. Pero de vez en cuando el papel arrugado caerá sobre la bandeja y, allí, inevitablemente, aparecerá un punto fijo. Es decir, de vez en cuando pensamos en nosotros mismos, y descubrimos nuestro punto fijo, el *yo*.

En la figura 94 hay un posible diagrama interno de un módulo predictor. La idea de este módulo es que a partir de una secuencia de datos de entrada y de salida se predice lo que va a llegar antes de que llegue. Entonces, si se retrasa la salida, debería coincidir con la entrada. Y si no coincide, un circuito optimizador debe corregir la predicción. Recordemos que en el cerebro tenemos muchas columnas corticales que actúan como módulos de predicción. Y que los *autoencoders* que se usan en *Deep Learning* hacen prácticamente lo mismo: predecir sus propias entradas. De modo que hay bastante coherencia en esta teoría.

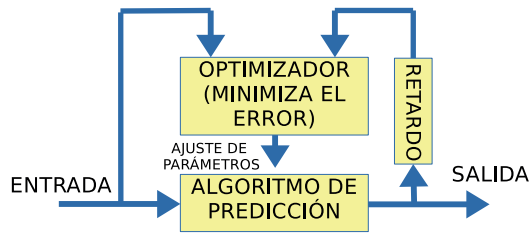


Figura 94. Módulo predictor

El *yo* es un invariante por el que tiene que pasar cualquier sistema inteligente que vaya a ser consciente. Para entender bien este concepto recomiendo volver a recordar el problema de ingenio de salvar el foso del capítulo anterior, junto con el punto por el que hay que pasar. El *yo* es un punto fijo también por esta razón.

Quiero también recalcar qué tipo de argumentación no es esta. No estoy diciendo que como los puntos fijos causan perplejidad, y como la consciencia también causa perplejidad, entonces la consciencia es un punto fijo. Lo que estoy diciendo es que hay un bucle cerrado de predicción que genera un punto fijo y que es sensato identificar ese punto fijo con el *yo*. Y que la noción del *yo* produce perplejidad como la producen todos los puntos fijos.

Evolutivamente hablando, si hay un sistema predictor (Figura 95), por mutaciones se puede cerrar el bucle, con lo cual emerge el concepto del *yo* (Figura 96).

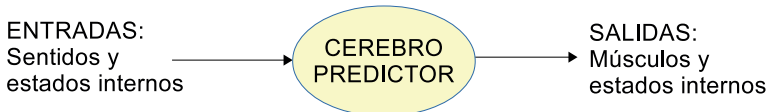


Figura 95. El cerebro, prediciendo lo que percibe y tomando acciones

La consciencia no termina aquí. Más bien, comienza aquí, con este bucle. Desde un punto de vista ingenieril, es este bucle el que crea la ilusión del *yo*, que se convierte en el sujeto de la narrativa de las sensaciones que recibo como entradas o como estados internos, y el aparente sujeto causal de las acciones que realizo. En ese primer nivel “predigo lo que voy a hacer y hago lo que estoy prediciendo”. Pero no puede olvidarse que el cerebro humano es muy complejo, de modo que es de esperar que haya muchos otros bucles actuando en varios niveles de descripción, en paralelo, y con unos sirviendo como metáfora, entradas o regulaciones de otros. Por ejemplo, al involucrar los sentidos “veo que mis manos se mueven como ordené, a la vez que hago mover las manos como lo predije”. Todos esos bucles hacen referencia implícita al *yo*. Crean el *yo*. Y ello construye nuestra rica experiencia consciente.

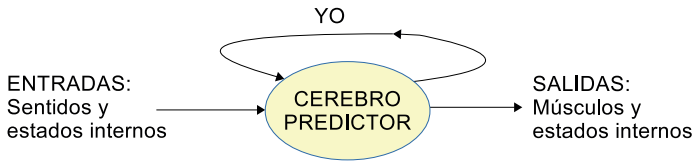


Figura 96. En el cerebro se forma un nuevo bucle

Además, si recordamos que las mentes *popperianas* contienen un modelo del mundo implementado como un simulador, desde esta perspectiva también emerge un punto fijo: dentro del mundo hay un objeto que simula el mundo. Se trata de otra función contractiva que se mapea sobre sí misma, que contendrá el mismo punto fijo: el yo.

Como mencionábamos en los experimentos al inicio de este capítulo, la consciencia es muchas veces un observador, y no quien dirige la acción. En ellos se ha detectado un retraso de tiempo positivo entre la acción y la consciencia de la acción. En este modelo que propongo, ello surge naturalmente. El *yo* es solo un punto fijo, sin existencia como objeto. Es la correspondencia entre lo que predigo que voy a hacer y lo que hace ese cuerpo, que resulta que es el mío. Los módulos de predicción vaticinan que va a ocurrir una acción en mi cuerpo porque las salidas neuronales para desencadenar esa acción ya han ocurrido. Por cuestiones de causalidad, esta predicción sucede unas décimas de segundo después de que la acción ya se haya decidido a nivel inconsciente. A su vez, como somos incapaces de detectar por nosotros mismos esa diferencia de tiempos tan pequeña, reinterpretamos la situación como que el *yo* ha decidido realizar una acción.

Desde mi punto de vista, eso no significa que sea imposible tomar decisiones conscientes. Lo que significa es que es imposible hacerlo en el mismo momento en que se debe reaccionar rápidamente. Pero cuando hay suficiente tiempo disponible, la consciencia sí puede tomar decisiones, básicamente creando situaciones reales o, la mayoría de las veces imaginarias (usando nuestro simulador *popperiano*), donde entrenar al inconsciente para que en un futuro enfrente esas situaciones de cierta manera.

De la misma forma, un guardameta de fútbol se entrena *offline*, que es cuando tiene suficiente tiempo, para crear situaciones que probablemente se van a dar durante un partido real, adiestrando así sus músculos, sus sensores y su sistema de decisión inconsciente. Ya que cuando tenga que detener un *penalty* real, la consciencia no va a tener tiempo para tomar decisiones elaboradas. Un ejemplo similar es cuando una joven decide entrar a la universidad. En un futuro, durante su vida profesional no va a tener tiempo de ponerse a estudiar usando libros para resolver cada problema que se le presente. Pero el entrenamiento previo ya lo hizo en las aulas, recabando infor-

mación y adiestrando habilidades. Es allí donde interviene la consciencia, en modo *offline* entrenando el inconsciente para situaciones futuras. Y ello tiene una consecuencia filosófica y jurídica: definitivamente sí somos responsables de nuestras decisiones, así sea porque no nos entrenamos para afrontar los problemas (estudiando o practicando), cuando había tiempo para ello.

El yo es solo un aspecto de la consciencia, pero está en el núcleo de todo lo demás. Con esta propuesta se abre un camino para que la inteligencia artificial tenga un yo, con todas las consecuencias positivas y negativas que pueda haber. Sin embargo, queda un obstáculo bastante filosófico y es que actualmente no disponemos de *tests* de consciencia. Hay muchos *tests* de inteligencia como IQ-test.cc (2017), 123test (2017), IQ_Research (2017) o Mensa (2017) que miden el IQ⁸⁸ de una u otra manera. Son bastante discutibles porque contienen sesgos culturales, idiomáticos y otros, pero algo indican. Sin embargo, no hay nada parecido para medir la consciencia. Y es importante tener *tests* de consciencia, para poder constatar los progresos que haga la inteligencia artificial en este sentido.

No obstante, hay un indicador de consciencia, quizás el único hasta ahora. Se trata de poner al sujeto bajo *test* delante de un espejo y mirar si se reconoce a sí mismo, o dicho en mis términos, si puede predecir lo que hace la imagen en el espejo con error igual a cero. Esta prueba la han pasado con éxito las siguientes especies: los grandes simios (humanos, chimpancés, gorilas, bonobos y orangutanes), los delfines, las orcas, los elefantes asiáticos, algunas hormigas y un pájaro llamado *magpie* según recogen Pachniewska (2017) y Think_Elephants_International (2015). No tengo constancia de que se haya hecho la prueba con pulpos ni con cuervos, que también tienen mucha inteligencia. Pensemos que aunque una especie haya pasado el *test* no significa que todos sus miembros lo hagan. Hay algunos individuos, típicamente los más fuertes, que simplemente muestran comportamientos agresivos frente al espejo.

La primera fase de un animal al verse ante un espejo es de comportamiento social. Creen estar ante otro individuo y reaccionan con amistad, agresividad o indiferencia. La mayoría de los animales como perros, gatos y peces se comportan agresivamente y nunca salen de esa fase. La segunda fase es verificar el comportamiento. Intentan distintas posturas y movimientos para comprobar si la imagen los imita. La tercera fase es de comprensión de que esa imagen es de uno mismo. Allí prueban a mirarse partes de su cuerpo que nunca antes habían podido ver, e incluso a coquetear delante del espejo. La cuarta fase la imponen los investigadores humanos para verificar

88 *Intelligence quotient.*

todo lo anterior. Para ello colocan una marca claramente visible en el animal. Pero el animal no la puede ver, excepto por medio del espejo. Cuando el animal la descubre, si se interesa por la marca en la imagen, ello indica que no ha comprendido nada. Solo si se interesa por la marca en el propio cuerpo es que ha pasado esta prueba de consciencia.

EL “HARD PROBLEM”

Con ello queda explicado el *yo* pero todavía hay un problema aparentemente más misterioso: la percepción subjetiva de la realidad, el flujo continuo de sensaciones que experimentamos, los *qualia* que memorizamos y recreamos en nuestro interior, en fin, lo que Chalmers llama el problema difícil⁸⁹ de la consciencia. Solo que él lo resuelve de una forma un tanto esotérica, suponiendo que todos los objetos del universo tienen consciencia y que ella es, por así decir, la quinta fuerza fundamental, en lo que se ha dado en llamar *pansiquismo*, también defendido por el filósofo Christof Koch (2017), que cree que la inteligencia y la consciencia son cosas completamente distintas. Muy al contrario, en este libro abogamos que la segunda es una consecuencia de la primera y que ambas son fenómenos computacionales. Chalmers también opina que es imposible que los computadores logren consciencia, porque sigue pensando que no es lo mismo una simulación de una tormenta, que una tormenta real. La última moja. Pero eso es absurdo por dos razones. Primera: para los seres que vivan en el mundo de la tormenta simulada, la lluvia simulada también moja. Segunda: si consideras que nuestro mundo es solo información, que se procesa de múltiples formas, y sabes que cualquier máquina universal de Turing puede simular a cualquier otra, entonces, no hay diferencia entre vida, inteligencia y consciencia reales versus simuladas.

Al contrario que ellos, Dennett niega que existan esos *qualia* y la experiencia subjetiva que conllevan. Aquí voy a proponer una solución a este dilema, muy cercana a la de Dennett, pero partiendo de bases meramente computacionales (García, 2018).

Para empezar debo decir que estoy de acuerdo en que la experiencia interna del color rojo o de sentir mucha hambre, no es lo mismo que si un robot en su *software* tiene un par de variables como:

color= RGB (255,0,0)

hambre= 95%

Ec. 38

No es lo mismo, pero... sí. En el fondo es lo mismo. Me explicaré.

89 *Hard problem.*

Sabemos que no hay un homúnculo⁹⁰ dentro del cerebro, o sea, no hay ningún control central. Por el contrario, sabemos que la información se encuentra allí distribuida en grandes redes neuronales. De modo que definitivamente no hay variables con valores asignados como en la ecuación 38. Pero nada impide que haya algo similar distribuido, usando algo así como memoria asociativa (parecida a la que se emplea en las *cachés* de los computadores). El problema no es la implementación, que puede ser cualquiera. Sea como sea, la información de la ecuación 38 está allí y los neurocientíficos no tardarán mucho en explicarnos cómo se almacena en el cerebro.

Pensemos primero si los robots van a tener *qualia*. Un robot como los que fabricamos actualmente va a tener variables como las de la ecuación 38 porque emplearemos algún lenguaje de programación estándar, y todos ellos tienen variables, pues es la forma más obvia de representar información. Y si a un robot le preguntamos por su estado interno, por sus *qualia*, responderá con toda precisión “estoy viendo que, al esconderse el sol, se produce un fenómeno de difracción de luz en la atmósfera cuyo color predominante se representa con el código RGB 255,0,0 que cae en la franja de lo que vosotros humanos llamáis rojo; y mis baterías se han agotado en un 95% por lo que necesito recargarlas” al más puro estilo de Mr. Spock de *Viaje a las estrellas* (Roddenberry, 1966) o de Sheldon Cooper en *The Big Bang Theory* (Lorre y Prady, 2007). Desde luego que no está sintiendo hambre y no está experimentando el esplendoroso atardecer de la figura 97, sino que solo puede dar detalles de sus estados internos de manera fría y objetiva.

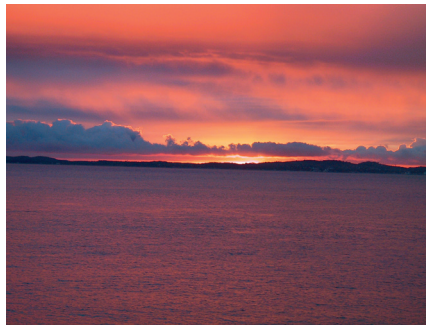


Figura 97. Atardecer en Salvador, Bahía

90 Si crees que dentro de tu cerebro hay un pequeño personaje tomando tus decisiones, entonces dentro del cerebro de ese personaje debe haber otro, y luego otro y otro... lo cual no tiene mucho sentido. El cerebro es un sistema distribuido de neuronas sin control central, del mismo modo que veíamos que los hormigueros y bandadas de aves no tienen un jefe que las guíe. En las personas más extrovertidas, una parte del sistema distribuido es externo al individuo, que no puede tomar decisiones o ni siquiera existir con cordura sin contar con la red de apoyo de sus amigos.

Hacer que el robot no solo pueda reportar el valor de descarga de sus baterías, sino que también sienta realmente hambre es bastante sencillo. Teniendo en cuenta, como dijimos, que fue creado por medio de un proceso evolutivo, eso significa que él sobrevivió, mientras que todos aquellos robots que no supieron buscar una fuente de energía cuando sus baterías estaban próximas a descargarse murieron en el proceso y no dejaron descendientes. Por ello, es de esperar que este robot cuando vea una fuente de energía se detenga a observarla, a sopesar cual es el nivel de batería que todavía dispone, a calcular si es suficiente y si puede llegar a otra fuente de energía. Conforme la batería esté baja, el robot comenzará a distraerse de sus tareas habituales, buscando con afán una fuente de energía e incluso corriendo hacia ella cuando su propio nivel esté en un punto crítico, olvidando sus otras obligaciones. Sus ojos se le irán hacia allá y todos sus ademanes delatarían la tensión interna de vida o muerte a la que se encuentra sometido. Cualquier observador diría que está sintiendo hambre, pues muestra todos los síntomas. La evolución hizo el trabajo de cablear en *hardware*, por motivos de supervivencia, unas respuestas que muestran intencionalidad⁹¹, ganas de acercarse al enchufe, vaya, está sintiendo hambre de energía. Si quieres saber cómo la evolución dota de intencionalidad a un robot, puedes ver el video de Karl Sims (1994), especialmente en la sección donde el objetivo es seguir un punto rojo. Se trata de un robot que aprende a nadar y a seguir un punto rojo por medio de algoritmos evolutivos, después de muchas generaciones y sin que ningún programador humano haya escrito ni una línea de código para dirigir su comportamiento. Todo fue aprendido en la larga lucha por sobrevivir, pues a los robots que lograban acercarse más rápidamente al punto rojo se les permitía tener más hijos. Al ver este video, muchos de mis estudiantes dicen que al robot *le da rabia* cuando el punto rojo se aleja, y vuelve a nadar vigorosamente hacia él. Dicen que el robot tiene un propósito claro, que tiene intenciones de seguir al punto rojo a donde quiera que vaya. Podemos decir que su estado interno, la percepción de que el punto rojo se aleja de él, le suscita emociones imposibles de controlar, que le empujan a seguirlo. La evolución cablea⁹² esas emociones.

Volviendo al ejemplo del robot hambriento (basta con interpretar el punto rojo como una fuente de energía) es importante entender esto en toda su profundidad: si le preguntamos al robot por qué se afana en aproximarse a la fuente de energía en vez de continuar el trabajo que estaba realizando nos

91 Intencionalidad en el sentido de mostrar intenciones de hacer algo. La definición filosófica de intencionalidad abarca este concepto pero es bastante más compleja.

92 Cablear es un concepto electrónico que significa que no es programable, que nacimos así con ello.

dirá que no lo sabe, que es algo más fuerte que él, que no lo puede controlar. Y ello es debido a que los comportamientos encaminados a lograr la supervivencia tienen prioridad sobre cualquier otro comportamiento. Los robots donde esto no es así hace tiempo dejaron de existir, por razones evolutivas.

El problema grave está con los *qualia* que no producen efectos observables, como regocijarse ante una puesta de sol. La evolución aquí no presiona en ningún sentido pues no se trata de una cuestión de supervivencia. Quiero matizar esto. Desde luego que poder diferenciar colores sí es importante para sobrevivir porque permite distinguir mejor a animales peligrosos, pero que esa combinación de colores me produzca más emoción estética que aquella no sirve para nada práctico. No tiene ningún valor evolutivo orientado a la supervivencia o reproducción.

Entonces aunque la representación de los colores en nuestra mente si puede ser el resultado de un proceso evolutivo, el placer estético que sentimos al verlos o recuperarlos de nuestras memorias no equivale a un frío código 255,0,0. La sensación subjetiva del color y, en general, de todos los *qualia* que no producen efectos observables, no tiene una explicación evolutiva.

La explicación de esos *qualia* viene de otro lado: se trata únicamente de un problema de codificación. Los *qualia* contienen información sensorial del mundo y de nuestros estados internos, pero esa información está encriptada en cierto modo. Podemos recuperarla, podemos trabajar con ella, pero no podemos entenderla en su detalle.

Ello se debe a que el *hardware* de un robot es muy distinto al *wetware* de los cerebros biológicos. En los robots la información está codificada en binario y desde allí se puede traducir biunívocamente a cualquier otro código numérico que represente números en decimal, letras, colores y sonidos. Pero la representación básica es binario. Mientras que en los cerebros biológicos la información básica es difusa y está representada por disparos eléctricos de neuronas. Y cualquier otra información (sonidos, letras, colores...) se soporta directamente en ese código, como conjuntos y secuencias de disparos eléctricos, a lo que se le suele llamar patrones de actividad neuronal. Además, las representaciones no son únicas y esos patrones se intersecan unos con otros a la vez que están distribuidos por amplios conjuntos de neuronas. Es decir, una neurona que dispara de cierta forma cuando la consciencia experimenta el color rojo, también puede estar disparándose de forma similar cuando la consciencia detecta un sonido particular. Los patrones son distintos y las experiencias que representan también, porque el conjunto de neuronas involucrado es distinto, pero puede haber algunas intersecciones.

Por ello no es de extrañar que ciertos colores o ciertos acordes menores nos pongan tristes: hay una intersección en los patrones de tristeza y de

música en acorde menor. Algo similar pasa con las representaciones computacionales en matrices dispersas que usa Hawkins en su cerebro artificial NuPic: hay una pequeña probabilidad de colisión entre símbolos distintos. Además, en el córtex una señal eléctrica que represente un sonido y otra que represente una imagen son indistinguibles por sí solas (Hawkins y Blackeslee, 2004). Son patrones que no están etiquetados. La diferencia está en el contexto, en las relaciones con otros patrones, de la misma manera que en un grafo dos nodos pueden parecer muy similares en cuanto a sus patrones de conexiones, pero los nodos vecinos son distintos y uno se refiere al sonido mientras que el otro lo hace a la visión.

En resumen, el *wetware* neuronal no es tan preciso como el *hardware* digital⁹³. ¿Por qué se dan esas interferencias y no otras? ¿Por qué al oír un acorde menor me pongo triste en vez de sentir hambre? Es algo accidental, aleatorio, que está embebido en nuestra arquitectura cerebral y que se irá descubriendo en detalle conforme se obtenga el mapa completo del cerebro. Pero es absolutamente accidental pues no ha habido allí presiones evolutivas para que ocurra de esa manera. En realidad sí las ha habido en el sentido de que las interferencias nunca son buenas. Pueden ser neutras (las que ocurren entre estados totalmente internos del cerebro, como música, color y tristeza) y pueden ser perjudiciales (las que producen comportamientos externos observables, como sería ponerse a correr al ver el color azul y paralizarse al ver el color amarillo). Las perjudiciales han tenido que sufrir una selección adversa y, por ello, no es de esperar que queden muchas ni en mucha intensidad, pues quien las porte sería presa fácil de manipulaciones que probablemente acaben con su vida, con lo cual no dejará muchos hijos que sigan portando tales anomalías.

Retomando el hilo, la codificación de los *qualia* es no numérica, distribuida, difusa y con una cierta ambigüedad. Está encriptada. Por “encriptada” no me refiero a que use una codificación fuerte tipo RSA de 1024 bits, sino simplemente que no hay una relación obvia, lineal y directa entre la entrada y la salida. Y tampoco hay una función inversa que me recupere una entrada a partir de una salida.

Analicemos un poco más los *qualia* del color. Podemos ignorar la intensidad de la luz, pues la verdad es que está codificada de una manera bastante lineal en nuestros *qualia*: más luz externa produce más sensación interna. Centrémonos únicamente en el color, que debería ser una única dimensión: la longitud de onda. Sin embargo, debido a que la evolución biológica es

93 Y, precisamente por ello, es más robusto el cerebro. Si unas neuronas se mueren no pasa nada, pero si un cable de computador se rompe, todo el computador deja de funcionar.

bastante chapucera⁹⁴ y optimiza solo en el corto plazo, los ojos humanos terminamos con tres sensores, cada uno de ellos centrado en una longitud de onda distinta, aproximadamente en el rojo, verde y azul. Los ojos entregan información de la diferencia de la longitud de onda de la luz incidente respecto al centro de la banda de cada sensor, y lo que debería ser una única dimensión ahora son tres. Ese mundo de colores en el que estamos sumergidos es solo una ilusión. Son *qualia* que codifican de una forma retorcida una información externa lineal.

Hemos diseñado las cámaras que usan nuestros robots repitiendo ese mismo esquema, y lo llamamos codificación RGB, que usa tres números. Pero nada impediría, en principio, hacer una cámara que entregue por cada pixel un único número: la longitud de onda incidente, en nanómetros. Al respecto, podemos ver una tabla de traducción con algunos códigos en la figura 98.

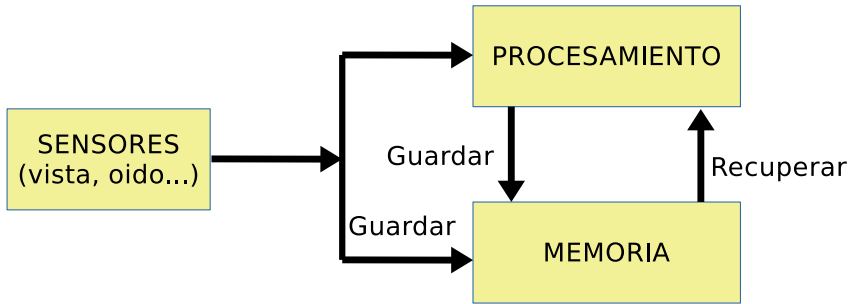
Longitud de onda (nm)	Código RGB	Nombre
380	97,0,97	Morado
440	0,0,255	Azul
486	0,239,255	Azul aguamarina
558	189,255,0	Verde limón
580	255,255,0	Amarillo
614	255,141,0	Anaranjado
699	255,0,0	Rojo
764	132,0,0	Rojo granate

Figura 98. Longitud de onda de la luz, código RGB y nombre convenido.
Adaptado de García (2018)

Si construimos un robot con una cámara así, y le pido que mire una manzana cuyo verde corresponde a 558 nm, y mañana le pregunto de qué color era, irá a su memoria, recuperará el color almacenado y me dirá que es 558 nm. Almacenó el número 558 y recuperó el número 558 (en la figura 99 podemos ver el correspondiente diagrama de bloques supersimplificado). Le pregunto después si su experiencia con los colores es tan maravillosa como la mía. El robot no lo puede saber, porque mis *qualia* de color permanecen

⁹⁴ El *bricoleur* de Jacobs.

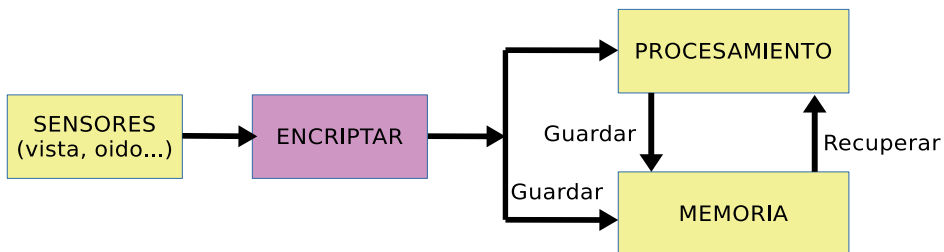
confinados en mi mente. Pero me dirá que su experiencia de color no tiene nada de maravillosa sino que se rige por la lógica: almacenó 558 y recuperó 558. El robot no disfruta de *qualia* de color en el mismo sentido que tenemos los humanos.



*Figura 99. Bloques internos de un robot sin qualia
Adaptado de García (2018).*

Es más, al ser tan precisa su representación interna de datos, puede externalizarlos sin que se pierda nada. El robot si tiene función inversa para cada *quale*. Y puede compararlos con los de otros robots, cosa que nosotros los humanos no podemos. Por ejemplo, no podemos llegar a determinar si el rojo que yo vi es más bonito que el que tú viste, y la razón de ello es que esa encriptación de datos nos lo impide.

Ahora hagamos algo maquiavélico con el pobre robot. Cuando esté dormido y sin que se dé cuenta, añadamos un módulo de encriptación de la información, justo antes de que se procese o almacene en memoria, de la forma que se indica en la figura 100.



*Figura 100. Bloques internos de un robot con qualia
Adaptado de García (2018).*

Tomemos también todos los datos previamente almacenados en su memoria (de experiencias que haya tenido en el pasado) y encriptémoslos, para que no pueda notar discrepancias con la forma antigua de codificar,

respecto a esta nueva. Cuando vea una pera verde y la compare con otra pera verde que vio hace unos días y que mantiene en memoria, ambas coincidirán aproximadamente. Pero ahora ya no puede hacer comparaciones con los colores que están viendo otros robots, pues ya no tiene función inversa para traducir un estado interno de su memoria a un valor sensorial externo. Para hacer el símil más exacto, deberíamos encriptar usando un código no numérico, quizás como un conjunto borroso de trazos (Figura 101) o algo similar. Entonces el robot se sentirá aturdido pues ve un color, en su memoria tiene ese color, pero no es capaz de expresar hacia afuera qué es y mucho menos compararlo con el de otros robots. El robot no podrá por menos que maravillarse, y hacer conjeturas sobre esto nuevo que está experimentando. ¡Acabamos de implementar *qualia* humanos en nuestro robot!

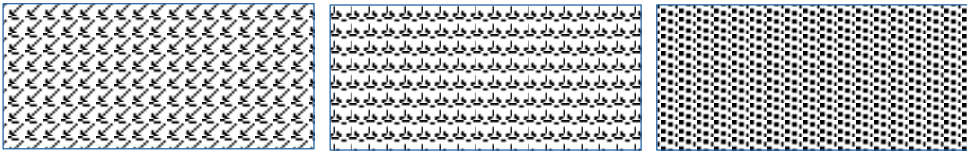


Figura 101. Patrones de información distribuidos, no comparables entre sí

Mi propuesta es que los *qualia* en humanos y la experiencia subjetiva que emana de allí son solo eso: datos codificados de una manera no numérica, no lineal y no invertible. Datos encriptados. Y a eso es lo que llamamos experiencia subjetiva, porque no se puede traducir a nada conocido⁹⁵. Y, sabiéndolo así, nos damos cuenta que los *qualia* no tienen mayor trascendencia, ni sirven para nada práctico. Es más, si hubiera suficiente presión selectiva, la evolución se encargaría de eliminarlos.

En el caso de los humanos, la encriptación ocurre directamente en el sensor. Por ejemplo, nuestros ojos nos dan la información en un formato raro de 3 variables (RGB) y además por medio de patrones de pulsos eléctricos.

Llamamos “sensación subjetiva” a aquella que no podemos objetivar, es decir, no podemos sacarla hacia afuera, para que la vean los demás. La “sensación de amarillo” entra encriptada en nuestra mente, donde se representa por un código único que es imposible volver a convertir hacia el exterior. Es decir, no tiene función inversa y por ello los códigos únicos se nos antojan tan misteriosos.

No es solo que la información va perdiendo detalles y se hace más abstracta conforme atraviesa más capas neuronales de nuestro cerebro. Es que

⁹⁵ Esto es a lo que David Chalmers llama el doble aspecto de la información, solo que mi propuesta lo despoja de misterio y lo plantea en términos computacionales.

realmente no hay como saber de dónde vino la sensación inicial. Hemos visto tres árboles, y podemos meditar con el número tres, que es una abstracción, y así pensamos en que es un número impar y primo. Pero más adelante podemos volver a imaginar los tres árboles sin ningún esfuerzo. La información original es recuperable. El problema con los colores es distinto. Hay una pérdida irrecuperable de información que nos impide volver atrás. El mismo color verde de los árboles lo podemos recrear en nuestra imaginación, lo podemos pintar en un cuadro y ajustar la paleta hasta que sea muy parecido al que tenemos en la memoria. Pero no sabemos de dónde viene ese color. No tenemos la noción de longitud de onda en ninguna parte de nuestra imbricada red neuronal. Y tampoco tenemos ninguna otra noción linealmente relacionada con la longitud de onda. La longitud de onda es un escalar mientras que el color es un vector de tres dimensiones dentro de nuestro cerebro, y la relación entre ambos es no lineal e incluso con ambigüedades. Eso es lo que nos impide entender el color.

También podemos plantear un experimento mental inverso al del robot: eliminar la magia de los *qualia* en los humanos. Para ello, con ayuda de equipo externo informamos a una persona de las longitudes de onda que está viendo en cada momento, y le pedimos que las memorice repitiendo el proceso muchas veces. Aunque a bajo nivel (neuronal) no existen los números, a alto nivel (simbólico) si los podemos manipular y memorizar en nuestro cerebro. De modo que el sujeto bajo pruebas, a la vez que memoriza el azul de la forma habitual, también memoriza 440 nm. Después de entrenarlo un rato con muchos colores le podemos mostrar un color concreto y pedirle que nos diga cuál es su longitud de onda. Lo hará sin sentir ningún misterio acerca de *qualia*.

Por último, pensemos en otro tipo de información interna, relacionada con la visión: gracias principalmente a que tenemos dos ojos, podemos triangular y averiguar la distancia aproximada a que se encuentra cada objeto. Es información interna, son también *qualia*, pero nadie se maravilla de ellos como ocurre con el color. Nadie dice “¡Uy!, qué distancia lejana tan hermosa” o “¡qué distancia intermedia tan especial!” mientras que sí lo hacemos con los colores. ¿Por qué? La respuesta es sencilla y viene a corroborar lo dicho hasta ahora: en las distancias sí ha habido presión evolutiva para que la codificación sea lineal, con el objetivo de permitir comparaciones. Es muy importante, por razones de supervivencia, saber si el tigre que se aproxima está muy cerca o muy lejos, o si el árbol al que me voy a subir está más cerca que el tigre. Mientras que no tiene ninguna importancia saber si

las hojas del árbol son más amarillas que la piel del tigre⁹⁶. Como resultado, la codificación de colores está arbitrariamente encriptada mientras que la codificación de distancias no lo está.

El hecho de que los *qualia* mantengan información privada desconectada del mundo real, aunado a que hay interferencia entre *qualia*, todo ello genera confusiones que interpretamos de una forma poética. Tenemos experiencias subjetivas. Pero si somos más rigurosos, lo que esto nos dice es que la evolución no ha trabajado todavía suficiente, seguramente porque no tenía incentivos para hacerlo. Nuestros cerebros son poco evolucionados. Puede haber cerebros mejores.

Un simulador, como el que propone Dennet que tienen las mentes popperianas, es un objeto computacional. De hecho, su formalismo corresponde a un computador. Y la forma más efectiva de controlar un computador es por medio de un lenguaje de programación. De manera que propongo una nueva utilidad de los lenguajes humanos. No solo sirven para comunicarse unas personas con otras, sino también cada persona con su simulador, o dicho más directamente, cada persona consigo misma. Las presiones evolutivas empujaron a crear lenguajes por esta doble ganancia. Y con ello los humanos (y cualquier animal que vaya adquiriendo un lenguaje) vamos abandonando el mundo de las experiencias subjetivas para alcanzar un grado más alto de consciencia. Experimentamos el color amarillo, podemos recuperar una vívida impresión de él si así lo deseamos, pero la mayor parte del tiempo simplemente usamos como sustituto la palabra “amarillo”, porque es más concreta y flexible, se le pueden añadir otras palabras para modificarla y matizarla. El uso del lenguaje permite un control más preciso de nuestro respectivo simulador y también permite comunicarse con otras personas. En este sentido, los *qualia* y la subjetividad son lo contrario de la consciencia y se han interpretado mal hasta ahora.

Estoy recordando un momento muy gracioso de la película *Words and Pictures* (2013) donde están enfrentados un profesor de lenguaje y una profesora de pintura. La profesora de pintura alega que una imagen vale más que mil palabras. El profesor de lenguaje pide a sus estudiantes que expliquen visualmente algunas ideas, tales como “estar hambriento” o “me gustan”, lo cual los estudiantes consiguen a base de gestos y gruñidos. Pero luego les pide que hagan lo mismo con “reúne a la tribu esta tarde que quiero explicarles que ya casi no hay comida y que tendremos que decidir hacia

96 Obviamente, sí tiene importancia el color para el humano, si desea descubrir al tigre camuflado entre hojas amarillas. Pero la presión evolutiva empujaría en este caso a lograr una mejor resolución en el sensor de color, donde se distingan más matices de amarillo. Y no a linealizarlo.

dónde movernos”. Los estudiantes quedan perplejos, pues es imposible visualmente transmitir esa idea. Con ello queda demostrada la superioridad y riqueza del lenguaje frente a una mera impresión visual; de la consciencia soportada por el lenguaje frente a la mera experiencia subjetiva.

Al respecto recordemos que en el capítulo anterior se mencionaban dos métodos de resolver problemas que empleamos los humanos: el intuitivo, que se corresponde con el mundo de las experiencias subjetivas; y el razonamiento estructurado, que es el que necesita un lenguaje para poder ser llevado a cabo.

Para terminar, quiero comentar que un átomo no es consciente en algún mínimo grado, como indican Chalmers (1995) y Tononi (2015). El cerebro es una acumulación organizada de átomos, pero la consciencia no es una suma de pequeñas consciencias atómicas. Analicemos esto con atención. El peso del cerebro es simplemente la suma de los pesos de los átomos, mientras que la consciencia del cerebro emerge debido a la interacción de un enorme número de neuronas. Cuando un átomo recibe las vibraciones mecánicas de una música escrita en acordes menores no es que sienta una tristeza muy, muy, muy pequeñita. Es que no siente ninguna. El concepto de tristeza no existe en las partículas elementales. No ha emergido todavía. El fotón no tiene intencionalidad, ni sentiencia, ni posee *qualia*. A lo sumo, quizás tenga algo de libertad, en el sentido explicado en el capítulo del mismo nombre.

LA CONSCIENCIA ES LA BUROCRACIA DEL CEREBRO

No hagas hoy lo que puedas dejar para mañana, pues a lo mejor mañana ya no hay que hacerlo,

CLARA E. GARCÍA BAÑOS

Hay aspectos del *yo* que son bien conocidos y que comparto plenamente. En particular, el *yo* no decide nada. No es quien está a cargo del cerebro y del cuerpo del individuo. No es quien goza de libertad, como discutíamos hace un rato. Y ya hemos visto las interpretaciones computacionales. Pero no quiero terminar aquí, pues hay otra forma de ver la consciencia y que nos puede servir sobre todo para entender por qué surgió. Sabemos que es lenta. Solo observa. Si toma acciones no son inmediatas sino de largo plazo. Se parece mucho a la burocracia de cualquier organización, y vamos a ver sus semejanzas.

Primero comencemos explicando la ley de Parkinson, que es empírica y que fue formulada por un funcionario inglés del mismo nombre en 1955, en India, mientras intentaba comprender por qué en la Oficina de la Colonia

Británica requerían cada vez más empleados. Se enuncia de muchas maneras. Una es así: “cualquier trabajo ocupa todos los recursos disponibles”. Parkinson se dio cuenta que el personal crecía exponencialmente, a una tasa de alrededor de un 5% anual, independientemente de que se logaran mayores o menores producciones de té. Y las razones que esgrimió para ello eran dos, que las voy a formular en términos modernos: la primera es que cada funcionario a lo largo de los años se cansaba de su trabajo y entonces contrataba a otros dos, más jóvenes, para que lo hicieran. De este modo repartía el trabajo entre los dos, pero impidiendo la comunicación directa entre ellos, con el objetivo de mantenerse imprescindible y evitar que lo despidieran. Nadie sabía todo sobre su trabajo, y todas las comunicaciones y decisiones pasaban por él. La segunda razón es que la gerencia quería saber el estado de la organización, pidiendo informes a los empleados que a su vez, para poder escribirlos, pedían informes a otros empleados, generando más y más trabajo.

Con ello aumenta exponencialmente la burocracia: las unidades internas se ponen tareas unas a otras, aun cuando no produzcan nada hacia el exterior. Y contratan más personal para poder hacer esas tareas agobiantes. En este sentido, la emergencia de la consciencia parece inevitable una vez que tienes muchos agentes encargados de realizar tareas concretas, usando los datos generados por otros agentes o el entorno, y ofreciendo los resultados al entorno o a otros agentes.

En el cerebro algo está pasando que se parece bastante, pues aunque constituye el 2% del peso del cuerpo, consume aproximadamente el 20% de sus recursos energéticos. Recordemos además que la consciencia no toma decisiones sino que mira lo que ocurre y lo anota, y discute consigo misma sobre los porqués, y compara con situaciones pasadas y se deprime, y en la confusión decide no actuar cuando podría hacerlo. Como Aureliano Buendía⁹⁷, la consciencia registra y pone etiquetas a todo. Y como la oficina que investigó Parkinson, se inventa tareas que ponerse internamente, unos módulos a otros, sin ejecutar hacia afuera ninguna acción.

Por cierto que la ley de Parkinson tiene corolarios en muchos ámbitos, incluida la computación: “el sistema operativo usará todos los recursos disponibles”; o “los discos duros se llenan por muy grandes que sean”. Pero también tiene una aplicación bastante interesante para administrar el tiempo: “si lo dejas para el último minuto, te llevará solo un minuto hacerlo” y eso muestra que la procrastinación no es tan mala como todo el mundo cree,

97 Ver cita del principio del capítulo.

en consonancia con la cita del principio de este apartado. La procrastinación puede ser pereza, pero también puede ser una estrategia de optimización.

Hay otra conclusión similar a la de Parkinson, que se llama el principio de Peter: “cada funcionario asciende en su organización hasta alcanzar el máximo nivel de incompetencia”. Claro que no se puede generalizar, no son todos, pero sí es un fenómeno bien conocido, que los funcionarios ascienden hasta llegar a un puesto donde no entienden ni siquiera cuáles son sus objetivos. Los funcionarios que todavía no han alcanzado su nivel óptimo de incompetencia son los únicos que realizan un trabajo efectivo en una empresa. Es una realimentación negativa, que estabiliza a estos funcionarios en sus respectivos puestos donde ya no pueden ascender más. Y eso se nota especialmente en los comités de planeación y de control de calidad. A estos funcionarios, permítanme el chascarrillo, se les podría llamar *bi-reguladores*, porque regulan regular. Controlan lo suficiente para mantener el sistema en la mediocridad. Las personas creativas no suelen terminar en esos puestos porque para ellos es más importante hacer, que controlar lo que hacen los demás. Por otro lado, las personas con muy bajas capacidades no pueden alcanzar esos lugares. De modo que habitualmente es el sitio para los mediocres, en el sentido de que no son buenos, pero no son malos. Están en la mitad. Y puede que ello sea lo correcto socialmente. Colocar allí a gente muy creativa puede estresar demasiado a la organización. El control mediocre se encarga de que la organización funcione, de manera ineficiente, con inercia, con procesos que no están bien diseñados, con mucho desgaste personal, pero logrando que las tareas al final se hagan. Las personas sufren pero la organización simplemente funciona. Si el bosón de Higgs es el que otorga inercia a las partículas, diría de él que es un *bi-regulador*.

Y la conclusión que se puede sacar de todo ello es que la consciencia es lenta, ineficiente, se enreda, tiende a paralizar al organismo.

Para terminar, presento el objetivo de toda esta perorata: pensemos que si la consciencia es burocracia, quizás también debería considerarse que la burocracia es consciencia. En las organizaciones, el afán de interrogarse pidiendo informes, estandarizando comportamientos, especificando funciones, puestos de trabajo y procesos, es la forma que tiene el superorganismo de reflexionar sobre sí mismo, para proponer futuras mejoras. Cumple así con las funciones mínimas de la consciencia. ¿Qué le faltaría? Por un lado, que los informes se traduzcan en acciones para mejoras futuras, cosa que muchas veces no ocurre. Es decir, que se cierren bucles de realimentación. De nada sirve saber por ejemplo que un funcionario no trata bien al público si no se ponen luego mecanismos correctores. Por otro lado, el concepto del *yo*, para lo cual se requiere también que se cierren bucles y que en ellos apa-

rezcan módulos predictivos que intenten anticiparse a los resultados propios y de otros módulos, para que surja un punto fijo.

El surgimiento de la consciencia en el superorganismo lleva acarreada una enorme pérdida de libertad individual, como lo sabemos quiénes estamos ya sufriendo el proceso. Y las etapas iniciales suelen ser muy angustiosas porque los individuos perciben graves errores, típicamente en procesos que van a ser sistematizados con normas hasta volverlos inamovibles aun siendo ineficientes, e incluso existiendo oportunidades de mejora. Porque la consciencia no ve nada de eso, lo único que percibe es que el proceso sistematizado más o menos funciona. Los informes en formatos de papel dicen que todo va bien, aunque la realidad sea otra. A esto se le podría llamar disonancia cognitiva de las organizaciones y quizás conlleve en un futuro a procesos de mejora. Pero el inicio es lento y doloroso.

Y ese es el grave riesgo de la consciencia: su velocidad de adaptación a nuevas situaciones es baja, una vez que ha tomado decisiones respecto a situaciones antiguas. Pensemos por ejemplo que si queremos cambiar un proceso en una empresa, que podría tener un costo⁹⁸ aisladamente de $O(N)$, cuando hay normas, formatos y estándares que lo regulan el costo suele ser del orden $O(N^2)$ debido a las interacciones internas, o quizás más si ello implica cambiar otros procesos de los que depende. Pero lo mismo pasa con la consciencia humana: una vez se ha desarrollado un hábito, una vez que la consciencia ha decidido algo y entrenado al inconsciente para actuar de forma predeterminada, es muy difícil de cambiar. Esto es especialmente palpable en muchas personas mayores, que tienen un conjunto de creencias arraigadas imposibles de modificar aunque les convenzamos con razonamientos de que el mundo a su alrededor ya no es el mismo y se requieren cambios mentales para adaptarse a ello. Decimos que sus conexiones neuronales se han fosilizado. Desgraciadamente pasa incluso con algunos jóvenes, y desde luego, pasa con las organizaciones, lo que siempre me trae a la memoria la portada de uno de los primeros libros sobre sistemas operativos, que muestra a unos dinosaurios enormes atrapados en un pozo de brea.

Para terminar, miremos lo que está ocurriendo ahora mismo. La consciencia en los computadores está más cerca de lo que se piensa. Eso se puede ya palpar en los trabajos de Google (Simonite, 2017) donde para desarrollar los detalles complejos del *software* de inteligencia artificial se usa esa misma inteligencia artificial, lo cual genera un bucle de realimentación positiva, es decir, creativo, que conjeturo que dará lugar a la consciencia en breve.

98 Voy a emplear la notación de complejidad computacional: $O(N)$ significa lineal y $O(N^2)$ significa cuadrático.

Actualmente, a esa inteligencia artificial ya se le hacen preguntas eternas (como “¿cuál es el objetivo de la vida?”) y se están obteniendo respuestas que no son nada filosóficas (como “vivir para siempre”) sino el reflejo del sentir de la humanidad, lo cual es lógico puesto que los datos de entrenamiento proceden de conversaciones humanas. Falta muy poco para que emerja consciencia en las máquinas.

RESUMEN

Hemos dado una explicación a la consciencia, como fenómeno emergente y computacional. El halo de misterio se desvanece. La complejidad que requiere la consciencia es apenas un poco mayor que la de la inteligencia, por lo que es de esperar que todo ser más o menos inteligente tenga consciencia en algún grado, lo cual se ha corroborado en los mamíferos. La inteligencia proporciona capacidad de predicción habitualmente distribuida en módulos. Y la consciencia requiere que esos módulos estén conectados consigo mismos, para que emerja un punto fijo que es lo que llamamos *yo*.

Nos maravillamos de los *qualia* internos, como el color, pero no son más que una forma defectuosa de codificar la información. La información está encriptada porque no ha habido presión evolutiva para mejorar su codificación. En los *qualia* que producen efectos observables no hay esa encriptación y tampoco nos maravillan, porque sí ha habido presión evolutiva para mejorar la codificación.

La consciencia puede entenderse también como la burocracia del cerebro, que no toma decisiones, pero observa, evalúa, genera y pide informes. Al revés, la burocracia en las organizaciones sociales puede entenderse como la emergencia de la consciencia en estos superorganismos. Cuando ello ocurra, y seguramente ya está ocurriendo aunque no lo percibimos, la consciencia a nivel individual humano tenderá a desaparecer, pues entra en conflicto con el nivel superior. Para vivir en sociedad es mejor ser adaptativo, conformista y no ser crítico. Imagina si cada célula de tu cuerpo tuviera consciencia, libertad y una opinión particular sobre cómo hacer las cosas. Por ello es muy posible que a lo largo de generaciones la consciencia humana se vaya apagando. Y nadie se dará cuenta de ello, pues consta únicamente de fenómenos subjetivos. Solo al revisar textos antiguos, escritos antes del siglo XXI, la gente del futuro dirá que hablábamos de ciertos conceptos un tanto esotéricos, que quizás formasen parte del vocabulario de algún ritual mágico.

Por si eso fuera poco, la consciencia en los computadores está a punto de hacer su entrada en escena. ¿Estaremos preparados?

PARA SABER MÁS

Dado que esta presentación sobre la consciencia es un trabajo original mío, no puedo ofrecer recomendaciones de otros libros que hablen más sobre ello. No obstante, ya hay algunos científicos que imaginan que la consciencia es únicamente un patrón matemático, como podemos ver en el siguiente video:

- Max Tegmark (2014). *Consciousness is a mathematical pattern: Max Tegmark at TEDxCambridge 2014*. Recuperado el 16 de octubre de 2017. Disponible en: <https://goo.gl/fY8zdN>

Y desde luego hay excelentes libros que explican muchos más detalles sobre la consciencia, con experimentos de casos reales, como los siguientes:

- Oliver Sacks (1995). *Un antropólogo en Marte*. Barcelona: Editorial Anagrama.

En particular recomiendo este libro fascinante, donde Sacks cuenta muchos casos de autismo, cada uno distinto y de los cuales podemos aprender otra perspectiva de lo que significa “ser humanos”. El autismo no es una enfermedad. Y tampoco tiene una única caracterización sino que hay muchísimas formas de serlo, por lo que hoy se suele hablar del espectro autista o incluso de neurodiversidad. El cerebro tiene muchas arquitecturas alternativas y que la nuestra sea normal mientras que la de los autistas sea la alternativa es bastante discutible. El día que los autistas sean mayoría pensarán que nosotros somos las anomalías. Además, pienso que ser autista puede tener grandes ventajas adaptativas en el mundo que estamos fabricando, donde la consciencia de las máquinas y de los superorganismos (sociedades) está a punto de despegar.

- Daniel Dennett (1995). *La consciencia explicada: una teoría interdisciplinar*. Barcelona: Ediciones Paidós.

Otro libro excelente, donde se habla de la consciencia en todos los aspectos que puede interesar a un ingeniero. Aquí Dennett nos cuenta que no hay ningún homúnculo, que no hay un punto central donde confluyan todas las señales eléctricas de las neuronas y que la consciencia es un fenómeno distribuido por todo el cerebro. Nos explica muchos casos de daños cerebrales y lo que ello implica para la consciencia del sujeto, así como lo que de allí se puede deducir para entender este tema. Mi pequeño aporte personal es anclar los dos fenómenos extraños de la consciencia en el plano computacional. Pero todo lo que había que analizar y decir sobre ella ya lo ha dicho Dennett en este excelente libro.

REFERENCIAS

Libros, artículos y enlaces web

- 123test (2017). *Free IQ test*. Recuperado el 11 de octubre de 2017. Disponible en: <https://goo.gl/Na4jRy>
- Academo (2017). *Wavelength to Colour Relationship*. Recuperado el 5 de febrero de 2017. Disponible en: <https://goo.gl/G6Yfhf>
- Bernardo, Á. (2013). *¿Qué secreto se esconde bajo la ilusión de la mano de goma?* Recuperado el 5 de julio de 2017. Disponible en: <https://goo.gl/3ZVta2>
- Botvinick, M. y Cohen, J. (1998). Rubber hands ‘feel’ touch that eyes see. *Nature* 391(756). DOI: <https://doi.org/10.1038/35784>
- Buller, D. L. (2005). *Adapting mind. Evolutionary Psychology and the Persistent Quest for Human Nature*. Cambridge: The MIT Press.
- Chalmers, D. J. (1995). Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies*, 2(3), pp. 200-219.
- Chalmers, D. (2014). *¿Cómo explicamos la consciencia?* TED. Recuperado el 18 de julio de 2017. Disponible en: <https://goo.gl/hfMVSD>
- Crick, F. y Koch, C. (1990). Towards a Neurobiological Theory of Consciousness. *Seminars in the Neurosciences*, 2, pp. 263-275.
- Damasio, A. (2006). *El error de Descartes*. Barcelona: Editorial Crítica.
- DeFelipe, J., Markram, H. y Rockland, K. S. (2012). The Neocortical Column. *Frontiers in Neuroanatomy*, 6(22). DOI: <https://doi.org/10.3389/fnana.2012.00022>
- Dennett, D. (1996). *Kinds of Minds: Toward an Understanding of Consciousness*. New York: Basic Books.
- _____. (2003). *Freedom Evolves*. New York: Penguin Books.
- de Waal, F. (2011). *Comportamiento moral en los animales*. TEDxPeachtree, filmed Nov 2011. Recuperado el 4 de julio de 2017. Disponible en: <https://goo.gl/ML-PZUd>
- Díaz, A. (2015). 35 fantásticas ilusiones ópticas que harán que replantees seriamente tu cordura. Recuperado el 3 de julio de 2017. Disponible en: <https://goo.gl/15UgXH>
- Duro, R. J., Becerra, J. A., Monroy, J. y Calvo, L. (2017). Multilevel Darwinist Brain: Context Nodes in a Network Memory Inspired Long Term Memory. *7th International Work-conference on the Interplay between Natural and Artificial Computation (IWINAC2017)*, A Coruña: Springer Lecture Notes in Computer Science series, 10337, pp. 22-31. DOI: https://doi.org/10.1007/978-3-319-59740-9_3
- Edelman, G. M. (1992). *Bright Air, Brilliant Fire: On the Matter of the Mind*. New York: Basic Books.

- García, A. (2017). Robot's and human's self: a computational perspective. *7th International Work-conference on the Interplay between Natural and Artificial Computation (IWINAC2017)*, A Coruña: Springer Lecture Notes in Computer Science series, 10337, pp. 3-9. DOI: https://doi.org/10.1007/978-3-319-59740-9_1
- _____. (2017b). A Computational Theory of Consciousness: Qualia and the Hard Problem. *Kybernetes*. DOI: <https://doi.org/10.1108/K-10-2017-0387>
- Google (2016). TensorFlow. Recuperado el 29 de noviembre de 2016. Disponible en: <https://goo.gl/DPXXgT>
- Hawkins, J. y Blakeslee, S. (2004). *On Intelligence. How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines*. New York: Henry Holt and Company.
- Hofstadter, D. R. (2007). *I Am a Strange Loop*. New York: Basic Books.
- Hood, B. (2012). *The Self Illusion: How Your Brain Creates You - TAM 2012*. James-RandiFoundation. Recuperado el 15 de julio de 2017. Disponible en: <https://goo.gl/kDYjHD>
- IQ-test.cc (2017). *Test de inteligencia*. Recuperado el 11 de octubre de 2017. Disponible en: <https://goo.gl/WTtxiQ>
- IQ_Research (2017). *IQ TEST*. Recuperado el 11 de octubre de 2017. Disponible en: <https://goo.gl/iJxA4m>
- Kauffman, S. A. (1993). *The Origins of Order*. New York: Oxford University Press.
- Koch, C. (2017). Consciousness in Biological and Artificial Brains. Santa Fé Institute. Recuperado el 18 de julio de 2017. Disponible en: <https://goo.gl/V7Hrb9>
- Kurzweil, R. (2013). *How to Create a Mind: The Secret of Human Thought Revealed*. London: Duckworth Overlook.
- Mensa (2017). *Mensa International*. Recuperado el 11 de octubre de 2017. Disponible en: <https://goo.gl/6Hquesu>
- Mikrojo (2016a). Ilusión Auditiva 'La Paradoja De Shepard'. Recuperado el 3 de julio de 2017. Disponible en: <https://goo.gl/KDPZ7g>
- _____. (2016b). *Ilusión Auditiva 2 'Phantom Words'*. Recuperado el 3 de julio de 2017. Disponible en: <https://goo.gl/YwchQf>
- _____. (2017a). *Ilusión Auditiva 3 'Sonido 3D (Hologonía)'*. Recuperado el 3 de julio de 2017. Disponible en: <https://goo.gl/xZzaXz>
- _____. (2017b). *Ilusión Auditiva 4 'El Efecto McGurk'*. Recuperado el 3 de julio de 2017. Disponible en: <https://goo.gl/wHEeAi>
- Narváez, S., García, A. y Gutiérrez, R. E. (2017). Execution of Written Tasks by a Biologically-Inspired Artificial Brain. *7th International Work-conference on the Interplay between Natural and Artificial Computation (IWINAC2017)*, A Coru-

- ña: Springer Lecture Notes in Computer Science series, 10337, pp. 45-52. DOI: https://doi.org/10.1007/978-3-319-59740-9_5
- Pachniewska, A. (2017). *List of Animals That Have Passed the Mirror Test*. Animal Cognition. Recuperado el 11 de octubre de 2017. Disponible en: <https://goo.gl/sev5Hh>
- Parkinson, C. N. (1955). *Parkinson's Law*. The Economist. November 19th, 1955. Recuperado el 18 de octubre de 2017. Disponible en: <https://goo.gl/WvyTva>
- Pinker, S. (1997). *Como a mente funciona*. São Paulo: Editora Schwarcz.
- Quoc, V. L., Ranzato, M. A., Monga, R. et al. (2012). Building High-level Features Using Large Scale Unsupervised Learning. *The 29th International Conference on Machine Learning – Conference Proceedings*, Edinburgh, pp. 127-137.
- Sacks, O. (2002). *El hombre que confundió a su mujer con un sombrero*. Barcelona: Editorial Anagrama.
- Salgado, R., Prieto, A., Bellas, F. y Duro, R. J. (2017). Motivational Engine for Cognitive Robotics in Non-Static Tasks. *7th International Work-conference on the Interplay between Natural and Artificial Computation (IWINAC2017)*, A Coruña: Springer Lecture Notes in Computer Science series, 10337, pp. 32-42. DOI: https://doi.org/10.1007/978-3-319-59740-9_4
- Sigman, M. (2016). *Your words may predict your future mental health*. TED. Recuperado el 6 de agosto de 2017. Disponible en: <https://goo.gl/bwKB5y>
- Simonite, T. (2017). Why Google's CEO Is Excited About Automating Artificial Intelligence. Recuperado el 18 de julio de 2017. Disponible en: <https://goo.gl/8Bz4NV>
- Sims, K. (1994). Evolved Virtual Creatures, Evolution Simulation. Recuperado el 17 de marzo de 2017. Disponible en: <https://goo.gl/RNDLtC>
- Spinney, L. (2015). The Time Illusion: How your Brain Creates Now. *New Scientist*. Recuperado el 8 de agosto de 2017. Disponible en: <https://goo.gl/mbSjb3>
- Think_Elephants_International (2015). *Mirror Self-Recognition in Asian Elephants!* Recuperado el 11 de octubre de 2017. Disponible en: <https://goo.gl/WYF1EV>
- Tononi, G. y Koch, C. (2015). Consciousness: here, there and everywhere? Royal Society. DOI: <https://doi.org/10.1098/rstb.2014.0167>
- van den Boogaard, E., Treur, J. y Turpijn, M. (2017). A Neurologically Inspired Network Model for Graziano's Attention Schema Theory for Consciousness. *7th International Work-conference on the Interplay between Natural and Artificial Computation (IWINAC2017)*, A Coruña: Springer Lecture Notes in Computer Science series, 10337, pp. 10-21. DOI: https://doi.org/10.1007/978-3-319-59740-9_2

-
- Vsauce (2017). Isolation - Mind Field (Ep 1). Recuperado el 7 de julio de 2017.
Disponible en: <https://goo.gl/KG2Z9T>
- Whitby, B. (1996). *Reflections on Artificial Intelligence*. Exeter, U. K.: Intellect Books.
- Wikipedia (2016a). *Banach Contractive Maps*. Recuperado el 14 de diciembre de 2016. Disponible en: <https://goo.gl/iA5Psi>
- ____ (2016b). *Brouwer fixed-point theorem*. Recuperado el 15 de septiembre de 2016.
Disponible en: <https://goo.gl/1rJ8bA>

Películas y videos

- Lorre, C. y Prady, B. (2007). *The Big Bang Theory*. USA: Warner Bros. Television.
- Roddenberry, G. (1966). *Star Trek*. USA: Desilu Productions & Paramount Television.
- Schepisi, F. (2013). *Words and Pictures*, USA: Roadside Attractions.

