

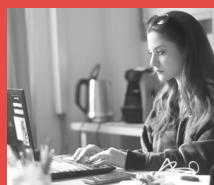
Bienvenidos a un viaje extraordinario a bordo de una carabela filosófica, guiados por el Autor, a través de “mares nunca antes navegados”. Esta Odisea por los mares tormentosos de la mente incluye contribuciones de 8 de los más brillantes expertos internacionales en el cerebro y la conciencia, incluido el premio Nobel de Física, Sir Roger Penrose. Dividido en dos partes, el libro comienza con Steven S. Gouveia delineando las coordenadas fundamentales de la filosofía y la ciencia de la mente. La segunda parte cobra vida en los diálogos del Autor con los 8 pensadores, brindando una experiencia única de exploración intelectual. Prepárate para zarpar hacia lo desconocido, donde la ciencia y la filosofía convergen en un viaje cautivador a través de los misterios de la mente consciente.



Steven S. Gouveia es licenciado en Filosofía y comenzó un Doctorado en Filosofía de la Mente y Neurociencia en la Universidad de Minho a los 22 años. A los 30, fue nombrado Profesor Honorario en la Facultad de Medicina de la Universidad Andrés Bello en Chile. Actualmente, es Investigador en el Instituto de Filosofía de la Universidad de Porto, liderando un proyecto sobre Ética de la Inteligencia Artificial en Medicina. Ha sido investigador visitante en la Universidad de Ottawa, Canadá, y ha publicado 14 libros académicos. Más info: www.stevensgouveia.weebly.com.



Mildreth Liliet Hernández Cruz. Licenciatura en Sociología. Maestría y Doctorado en Filosofía de la Ciencia por la Universidad Nacional Autónoma de México (UNAM). Estancia Posdoctoral en Ciencias de la Educación por la Universidad Autónoma del Estado de Hidalgo (UAEH). Profesora de la Facultad de Ciencias Políticas y Sociales (UNAM). Candidata al Sistema Nacional de Investigadores (SNI) en México.



Gaia Amoroso, estudiante ABAPA - Academia de Bellas Artes de Palermo (Italia), graduada en Graphic Design, especializada en gráfica editorial, ilustración y fotografía.

STEVEN S. GOUVEIA

LA ODISEA DE LA MENTE



STEVEN S. GOUVEIA

Susan Blackmore - Sir Roger Penrose
- Nicholas Humphrey - Karl Friston
- Joseph LeDoux - David Chalmers
- Christof Koch - Anil Seth

LA ODISEA DE LA MENTE

dialogos sobre el cerebro y la conciencia



LA ODISEA DE LA MENTE

Diálogos sobre el

CEREBRO

y la

CONCIENCIA

Steven S. Gouveia



2024

Steven S. Gouveia

MLAG | Instituto de Filosofía | Universidad de Porto, Portugal

ISBN: 9798336731521

© El Autor y colaboradores bajo licencia exclusiva.

Este trabajo esta protegido por derechos de autor y de traducción. Todos los derechos son únicos y exclusivamente licenciados por el autor y por los participantes, ya sea en su totalidad o en su parcialidad. Incluyendo los derechos de traducción, reimpresión, reutilización de imágenes e ilustraciones, recitación, difusión, reproducción de microfilmes o cualquier otra forma física. Así como transmisión, almacenamiento o recuperación de información, adaptación electrónica, software de computadora, por alguna metodología similar o diferente en la actualidad o desarrollada posteriormente.

El uso de nombres descriptivos generales, nombres registrados, marcas comerciales, marcas de tendencias, marcas de servicio, etcétera en esta publicación no implica, incluso en ausencia de una declaración específica, que tales nombres estén exentos de leyes y regulaciones de protección relevantes y, por lo tanto, estén libres para uso general. El autor y los participantes pueden asumir con seguridad que los consejos e información contenidos en este libro se consideran verdaderos y precisos en la fecha de publicación.

El autor como los colaboradores no ofrecen garantía alguna ya sea expresa o explícita respecto al material aquí contenido o por cualquier error u omisión que haya sido hecho. El Autor mantiene una postura neutral en lo que respecta a las reclamaciones jurisdiccionales en mapas publicados y afiliaciones institucionales.

Design and Portada: © Gaia Amoroso.

Caras originales de la Portada: © Ana Monteiro.

Traducción: © Mildreth Liliet Hernández Cruz.

Transcripciones: © Tássia Vianna.

Editor: Kindle Direct Publishing (Amazon).

Publicación original: "A Odisseia da Mente" (en portugués).

Índice

Prefacio _____ 7

Parte I | Conciencia

I. Introducción: Conciencia _____ 13

II. La Conciencia Misteriosa _____ 29

III. La Conciencia Alterada _____ 45

IV. La Conciencia Evolutiva _____ 63

V. La Conciencia Cuántica _____ 75

Diálogos I | Conciencia

VI. Diálogo con David Chalmers _____ 93

VII. Diálogo con Susan Blackmore _____ 109

VIII. Diálogo con Nicholas Humphrey _____ 131

IX. Diálogo con Sir Roger Penrose _____ 155

Parte II | Cerebro

I. Introducción: Cerebro _____ 179

II. El Cerebro Predictivo _____ 205

III. El Cerebro Libre _____ 219

IV. El Cerebro Integrado _____ 229

V. El Cerebro Emocional _____ 245

Diálogos II | Cerebro

VI. Diálogo con Anil Seth_____	261
VII. Diálogo con Karl Friston_____	281
VIII. Diálogo con Christof Koch_____	301
IX. Diálogo con Joseph LeDoux _____	319
Conclusión_____	339
Agradecimientos_____	343
Biografía del autor_____	349

Prefacio

El objetivo de este libro es llevar al lector a bordo de una carabela¹ filosófica, acompañado por algunos de los más grandes y valientes navegantes del siglo XXI, en una asombrosa expedición a través de los mares tempestuosos del conocimiento que abordan cuestiones complejas de la mente humana.

Esta Odisea de la mente nos llevará, inspirados por la valentía y coraje de intrépidos exploradores de los descubrimientos marítimos, a desafiar las olas y mareas desconocidas de la conciencia, en donde cada pensamiento es una estrella que nos guía en este cielo nocturno rodeado de misterios que, tal vez, nunca lograremos resolver.

¹ La carabela portuguesa se destacó por varias innovaciones, como el uso de la vela latina triangular que permitía una mayor capacidad para navegar contra el viento, así como su casco ligero y ágil que le otorgaba una alta versatilidad. Además, los portugueses también desarrollaron tecnología de navegación más avanzada en comparación con la utilizada hasta entonces, mediante el uso de la ballestilla y el astrolabio, lo que les brindaba la posibilidad de navegar en lugares del mundo que nunca antes habían sido explorados.

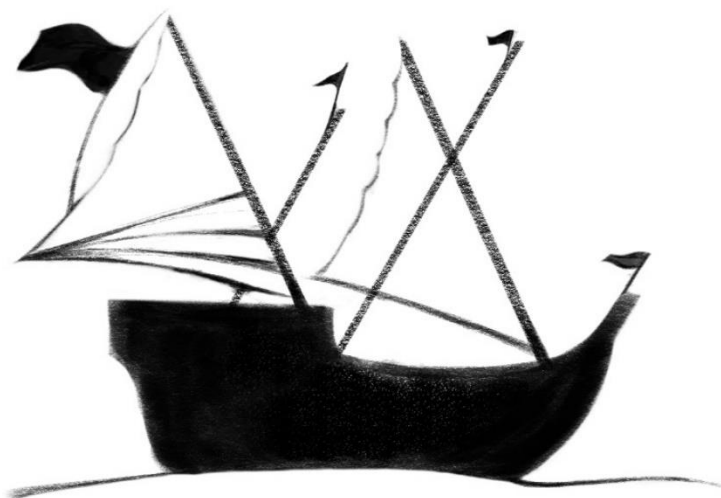


Imagen 1: "Caravella" © Gaia Amoroso

Al izar las velas de la curiosidad, seguramente encontraremos al menos una isla en donde podamos pausar en el viaje y descubrir nuevos colores, nuevas especies, más especias. Pero también, mayor deslumbramiento y curiosidad, y, por ende, más preguntas que responder.

Quizá cada sinapsis en nuestro cerebro pueda servir como un brújula -como la que está en el centro de la portada de este libro- que apunta hacia direcciones inexploradas y tierras prometedoras de mayor comprensión.

Esta incesante búsqueda -en esta gran Odisea de nuestro siglo- enfrentamos un gran enigma que permea los mares de la mente: ¿Qué le confiere a la mente su singularidad?

A medida que nos aventuramos entre arrecifes de pensamientos y bahías de reflexión, un mapa de la conciencia se va dibujando lentamente ante nuestros ojos.

Como los grandes navegadores que trazaron nuevas rutas por tierras desconocidas, enfrentamos las tormentas de incertidumbre y complejidad, y al mismo tiempo celebramos descubrimientos que amplían nuestro horizonte.

En cada página de este libro, encontraremos ilustraciones vivas de experiencias e ideas que espero inciten al lector a crear su propia carabela y explorar por sí mismo "mares nunca antes navegados".²

En última instancia, esta odisea no se trata solamente de desentrañar el misterio de la conciencia o la naturaleza del cerebro, sino más bien de abrazar el viaje hacia lo desconocido a través del astrolabio de la ciencia y la ballestilla de la filosofía.

La mente, como el océano, es vasta e inexplorada, y nuestra búsqueda es guiada por la pasión del descubrimiento y por la sed de comprender un misterio incesante que parece atormentar a todos los siglos de nuestra humanidad.

² Referencia a la primera estrofa del Canto I de la obra "Os Lusíadas" de Luís Vaz de Camões, poeta portugués que describió las aventuras de los descubrimientos marítimos por los portugueses.

Esperemos que al final de este viaje, hayamos conseguido encontrar un buen puerto, que nos permita por lo menos un breve y bien merecido descanso.

Para cerrar este prefacio, una nota menos poética. El presente libro tiene una estructura específica, se encuentra dividido en dos partes: la primera parte está dedicada a la conciencia y la segunda parte está dedicada al cerebro.

En cada parte el lector encontrará dos secciones. La segunda mostrará diálogos con 8 expertos de renombre internacional en los temas tratados, mientras que la primera incluirá introducciones temáticas. Estas introducciones tienen como objetivo ayudar a cada lector a apreciar los diálogos con mayor facilidad, a pesar de la complejidad de los temas presentados en este libro.

De tal forma que comience esta odisea de la mente.

Steven S. Gouveia
Porto, Portugal
10 | 10 | 2024

PARTE I

Conciencia

PARTE I

Conciencia

I. INTRODUCCIÓN: CONCIENCIA

Imagina que el lector se encuentra asimismo, en este preciso momento, en una de las más increíbles playas en el planeta tierra, como puede ser la maravillosa playa de Copacabana en Río de Janeiro, en Brasil.

Ahora imagina que estás disfrutando de una sabrosa caipirinha, la famosa bebida brasileña, mientras disfrutas del día soleado y de una hermosa vista del "Pan de Azúcar".

Mientras esta experiencia es probablemente mágica, para la mayoría de nosotros, hay otra magia ocurriendo en ese momento ¿Qué podría ser? La magia de estar saboreando una bebida específica, experimentando alegría y felicidad, todo ello sucediendo desde un punto de vista muy específico: ¡TU punto de vista!

Esta es la parte mágica de la historia: parece que nadie puede realmente explicar cómo esas experiencias conscientes están sucediendo en tu mente. Esto es uno de los más grandes misterios en el universo.: ¡Cómo una colección de neuronas inconscientes pueden dar lugar a un sentido unificado de experiencia subjetiva - una mente conciente-!

Debido a que es un misterio tan grande, académicos y pensadores de diversas disciplinas están empleando una variedad de herramientas y estrategias metodológicas para pintar un cuadro claro de cómo tu mente es moldeada por tu cerebro, tu cuerpo y tu conexión con el entorno (¡el mundo!).

La conciencia humana es, quizás, uno de los últimos misterios inexplicables del mundo en la actualidad. Por supuesto, no es el único: muchos otros misterios existieron en el pasado, como el misterio del origen del universo, el misterio de la vida y la reproducción, el tiempo, el espacio o la gravedad.



Imagen 2: "Architettura interna" © Gaia Amoroso

Aunque algunos de estos misterios pueden compartir la ausencia de una respuesta final, aún podemos contemplarlos. Estos misterios no han desaparecido, sino que han sido comprendidos porque supimos hacer las preguntas correctas.

Sin embargo, en lo que respecta a la conciencia, parece que no hemos llegado a ese punto: hay una profunda confusión, una intensa lucha de argumentos y contraargumentos, y diversas teorías intentando explicar el mismo fenómeno.

Recientemente, tuve el privilegio de presentar una ponencia aceptada por segunda vez en una de las conferencias académicas más importantes del mundo sobre la conciencia: la Conferencia de la Ciencia de la Conciencia.

Después de haber participado previamente en la edición de 2019 en el sublime Interlaken de Suiza, esta segunda participación ocurrió en mayo de 2023, en Taormina, una pequeña y maravillosa ciudad en Sicilia, Italia, con vistas al volcán Etna.

Como ponente en esta conferencia, en el primer día, recibí un código específico y un enlace web donde pude votar por la teoría de la conciencia que creía más probablemente correcta.

El lector quedará asombrado al descubrir que, al abrir la página en línea, me encontré con más de 13 (!) opciones para elegir; más de 13 teorías de la conciencia

estaban disponibles para seleccionar. Como puedes imaginar, si existen 13 teorías para explicar un fenómeno particular en el mundo, algo definitivamente debe estar equivocado.

Entonces, ¿cuál es el misterio de la conciencia? ¿Por qué parece tan desafiante, tanto para la ciencia como para la filosofía, formular una teoría plausible para explicar su existencia?

Todos tenemos experiencias subjetivas: percepciones, sensaciones, dolores e ideas. ¿Cómo pueden los cuerpos físicos vivos en un mundo físico producir tal fenómeno? Estas oscuridades vinculadas a la conciencia la convierten en uno de los problemas más excepcionales e importantes de la era contemporánea.

Algunos pensadores creen que la conciencia puede no ser tan misteriosa como inicialmente parece. Sacando analogías de la historia de la ciencia, argumentan que quizás toda esta oscuridad es un error creado por nuestro lenguaje y la forma en que usamos nuestros conceptos mentales.³

Por ejemplo, si observamos la forma en que el concepto de "luz" ha sido considerado a lo largo de la historia, quizás podamos aprender algo relevante relacionado con la conciencia.

³ Por ejemplo: Churchland, P. (1986) *Neurophilosophy: Toward a Unified Science of the Mind/Brain*, Cambridge: The MIT Press.

La teoría corpuscular de la luz, de origen griego y basada en el atomismo antiguo, describe la luz como compuesta por partículas específicas llamadas "corpúsculos".

Esta teoría fue defendida por ilustres pensadores, incluyendo a Sir Isaac Newton, quien argumentó que todos los procesos de reflexión y refracción de la luz solo podían explicarse si la luz estuviera compuesta por partículas. Esto se debía a que la alternativa, que la luz estuviera formada por ondas electromagnéticas, no podía explicar las trayectorias en línea recta.⁴

Interesantemente, este argumento tenía cierta base empírica: se sabe que Newton realizó alrededor de 40 experimentos que demostraron la naturaleza (supuestamente) corpuscular de la luz.⁵

Esta teoría fue considerada verdadera durante siglos, hasta que el científico Thomas Young, al concebir el famoso Experimento de la Doble Rendija en 1801, refutó la teoría corpuscular de la luz a favor de una teoría de ondas. En esta nueva perspectiva, la luz fue considerada tener una naturaleza similar a la del sonido, compuesta de ondas electromagnéticas.

⁴ Publicación original: Newton, I. (1704) *Opticks: or, A Treatise of the Reflexions, Refractions, Inflexions and Colours of Light*, London.

⁵ Publicación original: P. Rowlands, P. (2017) *Newton and Modern Physics*, London: World Scientific.

Esta teoría de las ondas podía explicar muchos fenómenos, como la difracción o la interferencia de la luz, que la teoría anterior no podía elucidar. Ahora, en este caso, el concepto inicial de "luz" en la teoría corpuscular fue, de cierta manera, reducido a otro concepto de "luz" - radiación electromagnética - de la teoría de las ondas.

Esta reducción tiene dos significados específicos: por un lado, la teoría corpuscular en su conjunto fue reducida a la teoría de las ondas; por otro lado, esta reducción fue tan exitosa que pudimos reducir el concepto de "luz" al concepto de radiación electromagnética.

¿Podría ser que solo encontremos este ejemplo de reducción en la historia de la ciencia, en el que creíamos que tendría perfecto sentido pensar en la existencia de ciertos conceptos, pero que, después de todo, descubrimos más tarde que eran seudocóncptos?

Según aquellos que creen que el concepto de luz en la teoría corpuscular está en el mismo nivel que el concepto de conciencia, hay más ejemplos que se pueden proporcionar.

La teoría del flogisto fue desarrollada dentro del campo de la química por Georg Stahl en el siglo XVIII. Stahl argumentaba que todos los cuerpos combustibles tendrían, en su composición, un elemento llamado

"flogisto" que se liberaba al aire durante los procesos de combustión.

Esta teoría fue una versión mejorada del mentor de Stahl, Johann Becher, quien publicó un libro llamado "Physica Subterranea", donde argumentaba que un elemento específico se liberaba cuando un material se quemaba.⁶

Es importante destacar que esta teoría del flogisto ya era una evolución de otra teoría antigua, la teoría de los elementos de Empédocles, que argumentaba que había cinco elementos en el mundo: fuego, tierra, aire, agua y éter.

Una vez más, esta teoría tendría una base empírica (ingenua). Stahl realizó una serie de observaciones en los procesos llevados a cabo en metalurgia: cuando un material se quemaba, sufría corrosión. Cuanta mayor fuera la corrosión, mayor sería la cantidad de flogisto liberada por ese material.

Esta teoría fue aceptada por los contemporáneos de Stahl debido a su mayor poder explicativo en comparación con la teoría de Empédocles. La teoría del flogisto permitió explicar por qué un material orgánico perdía masa cuando se quemaba: esto se atribuía a la pérdida del elemento flogisto hacia el aire.

⁶ Publicación original: Taylor, S. (2010) *Alchemists, Founders of Modern Chemistry*, Montana: Kessinger Publishing.

La imposibilidad de tener combustión sin aire también se explicaba por esta teoría: esto se debía a la imposibilidad de que el flogisto fuera capturado por el aire. Finalmente, el fin del proceso de combustión se explicaba por el agotamiento de la presencia de flogisto en el material.

Una vez más, tuvieron que pasar muchos años para que la teoría del flogisto -y con ello la relevancia del concepto de "flogisto" que se utilizaba para describir la realidad-, fueran cuestionados.

Uno de los problemas con esta teoría estaba relacionado con el hecho de que, a pesar de que la pérdida de masa fuera una realidad en los materiales orgánicos, lo mismo no sucedía con los metales, los cuales supuestamente, en procesos de calcinación (es decir, oxigenación), también deberían perder masa.

Sin embargo, esta predicción específica no se cumplió. Por el contrario, Antoine Lavoisier demostró que los metales, en el proceso de combustión, podían ganar peso, contradiciendo la tesis central de la teoría del flogisto.⁷

A pesar de esta evidencia contraria, fue necesario esperar hasta el descubrimiento (¡accidental!) del oxígeno por parte de Joseph Priestley en 1774. Se dio

⁷ Publicación original: Bell, M. (2005) *Lavoisier in the Year One*, New York: Atlas Books.

cuenta de que era este elemento químico el responsable de la combustión, en lugar del flogisto.⁸

Después de varios desarrollos en las décadas siguientes, la teoría química de los elementos se consolidó y terminó por reemplazar completamente la teoría de Stahl, que perdió toda relevancia en la descripción del mundo.

Una vez más nos enfrentamos con un concepto como el del "flogisto", el cual pareciera tener total relevancia. Sin embargo, la investigación podría demostrar posteriormente que dicho concepto fue un pseudoconcepto, un error lingüístico, que cuando ya no fue considerado, perdió toda su relevancia.

Continuando con la historia de la química, esta vez aplicada a la biología, podemos encontrar una teoría desarrollada en el siglo XVIII que argumentaba que el calor era un fluido invisible e inodoro que todos los cuerpos orgánicos poseían en su constitución.

Su cantidad fue directamente correlacionada con la temperatura de ese cuerpo: una mayor cantidad de fluido era equivalente a una temperatura más alta. Curiosamente, esta teoría fue propuesta por la misma persona que contribuyó a refutar el concepto anterior

⁸ Publicación original: Jackson, J. (2005) *A World on Fire: A Heretic, An Aristocrat and The Race to Discover Oxygen*, New York: Viking.

de “flogisto”: Lavoisier había reemplazado este concepto por otro, el de “fluido calórico”.

Dado que la teoría del flogisto era inconsistente con los resultados experimentales de Lavoisier, él propuso una alternativa conceptual: considerar este fluido calórico como la sustancia del calor.⁹

Según esta teoría, el fluido calórico tendría una existencia finita en su cantidad, transfiriéndose de cuerpos más calientes a cuerpos más fríos. Además, no podía ser creado ni destruido, por lo que la tesis central de esta teoría era su conversión constante, lo que explicaba las interacciones de temperatura entre diferentes cuerpos.

La incapacidad de la teoría calórica para explicar varios fenómenos, como la evaporación y la sublimación, propició la emergencia de una teoría más prometedora, la teoría cinética, a través del trabajo de Benjamin Thompson, también conocido como Conde Rumford.¹⁰

Rumford observó que el calor no era una sustancia material con una forma fluida, sino que más bien debería considerarse como energía en movimiento. A través de experimentos que demostraban que el calor

⁹ Publicación original: Fox, R. (1971) *The Caloric Theory of Gases*, Clarendon Press: Oxford.

¹⁰ Publicación original: Brown, G. (2001) *Count Rumford: The Extraordinary Life of a Scientific Genius – Scientist, Soldier, Statesman, Spy*, Gloucestershire: Sutton Publishing.

podía generarse por fricción, desafió la noción predominante de que el calor era un fluido indestructible.

Uno de sus experimentos más interesantes involucraba la fricción generada al perforar cañones, lo cual hizo notar que la cantidad de calor producida era mucho mayor de lo que la teoría calórica pudiera predecir.

Estas observaciones, entre otras, fueron fundamentales para la comprensión moderna del calor como una forma de energía que experimenta una transformación constante. Contribuyendo significativamente al desarrollo de la teoría de la conservación de la energía y, en el proceso, refutando por completo el concepto del fluido calórico.

De forma interesante, este tipo de error conceptual no solo está presente en la historia de la química: podemos observar un fenómeno similar en la historia de la psiquiatría. La teoría de la posesión demoníaca, durante siglos, sirvió como una explicación aceptada para numerosas enfermedades psiquiátricas.¹¹

La teoría de la posesión demoníaca era ampliamente aceptada durante la Edad Media, con diversas versiones surgiendo en diferentes periodos y regiones del mundo. Se utilizaba como explicación para

¹¹ Publicación original: Kemp, S. e Williams, K. (1987) "Demonic possession and mental disorder in medieval and early modern Europe", *Psychological Medicine*, 17 (1):21-9.

comportamientos considerados anormales según los estándares de ese tiempo.

El exorcismo a menudo era considerado el único tratamiento disponible. En los documentos históricos que han perdurado, podemos encontrar una relación directa entre tipos particulares de demonios y los síntomas asociados con enfermedades específicas, como son los trastornos psicóticos, neuróticos o la epilepsia.

Curiosamente, ciertos filósofos medievales que influyeron significativamente en otros ámbitos de la filosofía y la teología contribuyeron en gran medida a la prominencia de esta teoría.



Imagen 3: "I sensi" © Gaia Amoroso

Por ejemplo, Agustín de Hipona afirmó que los demonios poseerían un cuerpo material, mientras que Tomás de Aquino argumentaba que los demonios serían entidades no corpóreas, inteligibles y separadas del mundo físico.

Singularmente, la postura de Aquino fue respaldada por la noción de que numerosos fenómenos empíricos descritos en la literatura no podrían existir si los demonios realmente tuvieran una existencia corporal.

Esta teoría comenzó a perder relevancia con el avance de la psiquiatría como disciplina fundamentada en evidencia y teorías basadas en conocimientos médicos que evolucionaron en los siglos XX y XXI. Sin embargo, es importante señalar que, como se explora en algunos de los temas dentro de este libro, la naturaleza de las enfermedades psiquiátricas sigue presentando desafíos y preguntas significativas para científicos y filósofos.

El punto central es que como todos estos casos históricos ahora resaltados -la existencia del flogisto, el fluido calórico, los demonios- la sociedad ha llegado a reconocer que a través del avance del conocimiento, ahora esos conceptos se encuentran carentes de significados y relevancia para entender la realidad.

Este punto sirve para ilustrar una consecuencia crucial cuando nos damos cuenta de que, en última instancia, los conceptos empleados no son beneficiosos para

comprender el mundo: las mismas preguntas y problemas que dan forma al pensamiento de filósofos y científicos también se vuelven sin sentido y, como resultado, pierden la necesidad de una respuesta.

Para entender este punto esencial, examinemos la teoría geocéntrica, que afirmaba que el planeta tierra era el centro del universo. Según esta teoría, el sol y la luna orbitaban alrededor de la tierra una vez al día, mientras que las estrellas permanecían fijas dentro de lo que se llamaba la "esfera celeste", girando alrededor del eje de la tierra.

En este sistema geocéntrico, las grandes preguntas que todos los científicos y pensadores procuraban responder giraban en torno a las siguientes investigaciones: ¿Qué impulsa a la esfera celeste a girar alrededor de la Tierra? ¿Cómo están fijadas y adheridas las estrellas a esta esfera? ¿Por qué mecanismo la Luna y el Sol mantienen sus orbitas precisas alrededor de la tierra?

Todas estas preguntas constituían los principales desafíos con los que luchaban los pensadores de esa época, esforzándose por desentrañar y proporcionar explicaciones creíbles. Sin embargo, con el descrédito de este modelo, liderado inicialmente por Copérnico y posteriormente defendido por Galileo y Kepler, junto con el advenimiento de la mecánica clásica de Newton, estas preguntas perdieron todo significado.

Para el modelo heliocéntrico, la postulación de una "esfera celestial" era simplemente ilógica y absurda: lo que parecía requerir una respuesta o una teoría ya no necesitaba más explicación.

Algunos filósofos y científicos argumentan que ciertas preguntas sobre la conciencia pueden estar formuladas de manera que asumen una realidad significativa para conceptos que podrían ser fundamentalmente defectuosos o carentes de definiciones claras.

La comparación con teorías históricas que luego resultaron ser pseudconceptos plantea la posibilidad de que nuestras preguntas actuales sobre la conciencia necesiten una reevaluación en términos de su coherencia y supuestos subyacentes.

Esta es una provocación filosófica con la que tendremos que lidiar durante esta odisea de la mente. No podemos negar que todo lo que hacemos y sentimos es, de alguna manera, parte de la mente consciente.

Entonces podríamos preguntarnos a nosotros mismos: ¿Cómo es que sabemos tan poco sobre lo que es inherente a nuestra naturaleza humana? ¿Podría ser que la ciencia contemporánea lucha por proporcionar una explicación satisfactoria de la mente? ¿Existe una teoría de la conciencia lo suficientemente robusta como para dar cuenta de todos esos fenómenos?

En la primera parte del libro, intentaremos responder a estas preguntas a través de diálogos con cuatro de los

pensadores más relevantes sobre la naturaleza de la conciencia: David Chalmers, Susan Blackmore, Nicholas Humphrey y Sir Roger Penrose, con quienes tuve el honor de debatir.

Para contar con una mejor comprensión de los diálogos subsecuentes en la segunda parte de cada sección, en los siguientes capítulos se ofrecerá una breve introducción a los temas explorados en esas discusiones.

Comencemos por comprender por qué la conciencia posee esta aura enigmática en relación a su existencia y una capacidad de ofrecer explicaciones plausibles sobre sí misma.

II. LA CONCIENCIA MISTERIOSA

El primer diálogo cuenta con una participación del filósofo que colocó el tema de la conciencia en el mapa de los grandes debates filosóficos y científicos. Nos referimos a David Chalmers, profesor de Filosofía y Ciencia Neural, así como co-director del Centro de Mente, Cerebro y Conciencia en la Universidad de Nueva York. Es uno de los principales proponentes de la idea de que hay algo particularmente complejo sobre la naturaleza de la conciencia.

Como observarás en el debate, Chalmers introdujo el famoso "problema duro de la conciencia", aunque el problema no es del todo nuevo. En el fondo, es una característica fundamental de aquellos que aceptan el problema mente-cuerpo como una cuestión real, es decir, aquellos que creen que la mente tiene propiedades que no pueden ser agotadas por la fisicalidad del mundo.

Muchos fenómenos mentales como el aprendizaje, el razonamiento, la memorización, etcétera, pueden explicarse en términos de desempeñar cierto "rol funcional": si descubrimos la "función" de un cierto sistema, sabremos todo lo demás sobre ese sistema.

Sin embargo, para Chalmers, esto no es así en lo que respecta a la conciencia:

“Lo que hace un problema difícil *difficil* y casi único es que va más allá de los problemas sobre el desempeño de las funciones. Observé que incluso cuando nosotros hemos explicado el desempeño de todas las funciones cognitivas y de comportamiento en los alrededores de la experiencia -discriminación perceptual, categorización, acceso interno, informe verbal- todavía aún se mantenía una pregunta adicional sin respuesta: ¿Por qué el desempeño de estas funciones está acompañado por experiencia?”¹²

Si aceptamos que este problema es real, una de las aparentes consecuencias es que estamos defendiendo que una explicación científica de la conciencia realmente puede escaparse de nosotros.

En este sentido, otro filósofo esencial para comprender la naturaleza del problema "difícil" de la conciencia es el filósofo Thomas Nagel. Nagel percibe el problema como la "subjetividad" de los estados mentales conscientes.¹³

Este filósofo sostiene que los hechos sobre los estados conscientes son inherentemente subjetivos: solo pueden entenderse plenamente desde perspectivas

¹² Publicación original: Chalmers, D. (1996) *The Conscious Mind*, New York: Oxford University Press, p. 202.

¹³ Publicación original: Nagel, T. (1974) "What Is It Like to Be a Bat?", *The Philosophical Review*, 83 (4): 435-450.

limitadas, accesibles únicamente para el sujeto que experimenta esa experiencia consciente en particular.

Por ejemplo, en el escenario inicial de tomar una caipiriña en la playa de Copacabana, solo el lector podrá tener acceso directo y subjetivo a la experiencia del sabor de esa caipiriña, y nadie más podrá entender exactamente lo que el lector está sintiendo en ese momento.

Sin embargo, la explicación científica requiere una explicación objetiva de los hechos, que se aparte de cualquier punto de vista particular. Por lo tanto, los hechos sobre la conciencia eluden a la ciencia, haciendo que "el problema mente-cuerpo sea verdaderamente intratable".¹⁴

Chalmers acaba por introducir un argumento original e interesante llamado "El argumento del zombi filosófico" para demostrar esa característica de la conciencia.

Este argumento gira en torno a la posibilidad lógica de un mundo idéntico al nuestro, que carezca de cualidades secundarias (llamadas *qualia* en filosofía), y que esté desprovisto de cualquier estado de conciencia: un mundo zombi. En este mundo, existiría un equivalente zombi para cada uno de nosotros, idéntico a nivel molecular. Sin embargo, este zombi carecería de

¹⁴ Original publication: Nagel, T. (1974) "What Is It Like to Be a Bat?", *The Philosophical Review*, 83 (4): 435.

cualquier experiencia consciente, aunque, desde un punto de vista externo, parecería comportarse como si estuviera consciente.



Imagen 4: "Gemelli" © Gaia Amoroso

Considera el siguiente escenario: tú, el lector, te encuentras con tu celebridad favorita en la calle, alguien a quien has admirado desde la primera vez que lo viste en tu película favorita. Sin que lo sepas, esta persona es un zombi filosófico.

Con gran entusiasmo y en un momento de extremo romanticismo, le expresas tus sentimientos románticos a esta celebridad. Para tu sorpresa, la celebridad responde con igual entusiasmo, correspondiendo los mismos sentimientos, expresando amor y comportándose como si estuviera genuinamente enamorada/o de ti.

Este ejemplo, claramente ficcional, reconociendo la improbabilidad de que una celebridad famosa sea recíproca en tales sentimientos, la persona que interactúa con nosotros no siente de hecho nada, carece de sentimientos o de experiencia consciente. Desde la perspectiva del lector, esta persona parece tener conciencia solo por su forma de comportarse.

Siguiendo el resultado de este ejemplo hollywoodense, David Chalmers argumenta que podemos decir, indirectamente, que los zombis tienen una posibilidad lógica, aunque puedan no tener una posibilidad natural, si entendemos la organización funcional de las personas.

Por ejemplo, podemos concebir que un isomorfo de silicona del lector, con chips en lugar de neuronas, no desarrollaría conciencia, y que por lo tanto los hechos de nuestra organización funcional no necesariamente implican los hechos de nuestra conciencia.

En este punto, es importante destacar que si la noción de un "zombi filosófico" es conceptualmente coherente

y lógicamente posible, entonces, en ese caso, los estados conscientes y subjetivos no son idénticos a los estados físicos del cerebro.¹⁵

Esto nos lleva a otro filósofo, quien argumenta que la existencia de la conciencia en el mundo físico causa una "brecha explicativa". Joseph Levine sostiene que hay una "brecha explicativa" especial entre la conciencia y lo físico.¹⁶

El desafío de cerrar esa brecha explicativa es lo que da lugar, de cierta forma, al problema difícil de la conciencia. Levine argumenta que una buena explicación científica debe implicar deductivamente lo que explica, permitiéndonos inferir la presencia del fenómeno objetivo a partir de una demostración de leyes o mecanismos y condiciones iniciales.

La implicación deductiva es un tipo de relación lógica con la siguiente configuración formal: si las premisas de un argumento son verdaderas, entonces se sigue que su conclusión es necesariamente verdadera.

Tomemos en cuenta el siguiente ejemplo: dado que descubrimos que un relámpago no es más que una

¹⁵ Históricamente, el experimento mental de los zombis filosóficos tiene como propósito refutar la teoría fisicalista de la mente que sostiene que los estados conscientes pueden reducirse, en el mismo sentido que se describe en la sección anterior, a estados físicos específicos, en relaciones de identidad estricta.

¹⁶ Publicación original: Levine, J. (1983) "Materialism and qualia: The explanatory gap", *Pacific Philosophical Quarterly*, 64: 354-361.

descarga eléctrica, conociendo todas las condiciones adecuadas para una descarga eléctrica en la atmósfera en el momento t nos permite deducir que el relámpago debe haber ocurrido en el momento " t ".

Sin embargo, tanto para Levine como para Chalmers, este no parece ser el caso cuando pensamos en la conciencia: no importa cuán detallada sea nuestra especificación de los mecanismos cerebrales o las leyes físicas, siempre hay una pregunta abierta sobre si la conciencia está presente y acompaña a cierto mecanismo psicológico subyacente o no.

Por supuesto, muchos pensadores fisicalistas, que creen que todo lo que existe en el mundo tiene que ser explicado por las leyes de la física y la química, rechazan este argumento. Argumentan que un mundo en el que supiéramos todo sobre los mecanismos contendría necesariamente una explicación sobre la conciencia. Según esta perspectiva, la conciencia sería necesariamente generada a partir de cualquier conjunto de circunstancias físicas que desconocemos en el momento, pero que nada nos impide poder llegar a conocer en el futuro.

Para ayudar al lector a comprender mejor este "misterio de la conciencia", podemos buscar una valiosa ayuda del imaginario Dr. Mary, el neurocientífico más brillante que el mundo haya conocido (desafortunadamente, no tenemos este tipo

de conocimiento en el presente; esto es solo otro experimento mental).

Por lo tanto, Mary es una neurocientífica genial que sabe todo lo que hay que saber sobre las propiedades físicas del cerebro, especialmente aquellas relacionadas con la percepción visual del color. Además, Mary es tan brillante que también conoce todos los hechos físicos sobre la luz y los colores.

Por lo tanto, podemos afirmar que Mary sabe todo lo que hay que saber sobre la forma en que los seres humanos experimentan los colores visualmente, abarcando tanto los procesos cerebrales involucrados como los procesos físicos.

Sin embargo, hay un pequeño problema: Mary ha sido prisionera desde su nacimiento en una casa donde solo tiene acceso a su habitación. Peor aún, esta habitación no tiene ninguna luz; todo el conocimiento que Mary aprendió sobre el cerebro y la mente tuvo lugar en un entorno monocromático, utilizando materiales y herramientas en blanco y negro.

¿Qué busca mostrar este escenario sobre la conciencia? Este experimento mental tiene como objetivo resaltar la siguiente situación: imagina que Mary finalmente es liberada de esa habitación monocromática, y su novio, con quien intercambiaba cartas de amor durante años, finalmente la conoce en persona.

El novio, en un gesto romántico, le presenta a Mary una rosa roja: por primera vez, Mary se enfrenta a un objeto de color que nunca ha visto antes, a pesar de poseer todo el conocimiento teórico sobre el cerebro y los colores, incluidas las rosas rojas.

¿Mary ya sabría qué esperar al ver el color rojo de la flor por primera vez? ¿O aprendería algo nuevo en esta experiencia visual, algo que, después de todo, se le había escapado a pesar de poseer todo el conocimiento sobre cómo funcionan el cerebro y los colores?

Este experimento mental, originalmente desarrollado por el filósofo Frank Jackson,¹⁷ pretende ilustrar dos tesis particulares para el lector. En primer lugar, que las experiencias subjetivas vinculadas a las cualidades (*qualia*) tienen una existencia real, no siendo una mera fábula teórica de algunos filósofos y científicos. En segundo lugar, la tesis de que la mente consciente solo puede ser descrita apelando a propiedades físicas y químicas es errónea, dado que, en el escenario de Mary, hay verdades sobre los colores que parecen ser no físicas.

Aunque este argumento tiene alguna fuerza intuitiva, algunos pensadores¹⁸ creen que nosotros no podemos

¹⁷ Publicación original: Jackson, F. (1982) "Epiphenomenal Qualia", *Philosophical Quarterly*, 32 (127): 127-136.

¹⁸ Por ejemplo: Churchland, P. M. (1985) "Reduction, Qualia, and the Direct Introspection of Brain States", *The Journal of Philosophy*, 82 (1): 8-28.

conclusivamente derivar esas dos tesis de la experiencia de Mary. En el mejor de los casos, la conclusión contraintuitiva es que Mary ya sabría qué esperar al experimentar cuando su novio romántico le ofreciera la rosa roja.

Por supuesto, la aceptación de estas dos tesis dependerá, en esencia, de la posición filosófica en la que se encuentre el lector con respecto a la conciencia: si toma en serio la experiencia de Mary, es probable que crea que hay algo fundamental en la conciencia que no puede describirse adecuadamente mediante la ciencia actual.

Sin embargo, si crees que Mary ya sabe qué esperar cuando ve la rosa roja ofrecida por su novio, es probable que tengas la opinión de que no hay nada especialmente misterioso acerca de la conciencia.

Explorar el misterio de la conciencia en relación con los nuevos desarrollos tecnológicos y la posibilidad de que estemos viviendo en un mundo simulado, similar a la película "The Matrix", es un área relevante para investigar.

La posibilidad de que estemos viviendo en una simulación fue defendida por el filósofo Nick Bostrom en su popular artículo "¿Estás viviendo en una simulación computacional?"¹⁹ El argumento de la

¹⁹ Publicación original: Bostrom, N. (2023) "Are you living in a computer simulation?", *Philosophical Quarterly*, 53 (211): 243-255.

simulación presenta tres escenarios sobre el futuro del mundo:

- (i) Es muy probable que la especie humana se extinga antes de alcanzar una etapa "post-humana";
- (ii) Es sumamente improbable que una futura civilización poshumana pueda ejecutar un número significativo de simulaciones de su historia evolutiva (o variaciones de la misma);
- (iii) Casi con certeza estamos viviendo en una simulación computacional.

Bostrom argumenta que, entre estas tres proposiciones, al menos una debe ser verdadera. Supongamos que (i) es incorrecto: alguna civilización en el universo alcanza la madurez tecnológica.

Ahora, consideremos que (ii) también es incorrecto: ninguna civilización utiliza sus recursos para llevar a cabo simulaciones. Si es así, podemos (supuestamente) concluir lo siguiente: dado que estas antiguas civilizaciones tienen la capacidad de ejecutar un enorme número de simulaciones, si las dos primeras posibilidades son falsas, puede haber un mayor número de entidades simuladas que de entidades no simuladas en este universo.

En este escenario, todos los seres humanos estarían, al menos la mayoría, viviendo dentro de simulaciones y

no fuera de ellas. El razonamiento presupuesto en el argumento es que, si rechazamos las dos primeras hipótesis, tendremos que aceptar que la tercera sigue necesariamente, dado que, por cada mundo real, tendríamos millones de mundos simulados.

Por lo tanto, la probabilidad de que actualmente estemos en una simulación parece ser genuina, si aceptamos las premisas planteadas por Bostrom, algo que Chalmers parece aceptar, defendiendo el siguiente argumento:

(iv) Es muy probable que las simulaciones probables conscientes sean posibles;

(v) Es más probable que, si simulaciones humanas conscientes fueran posibles, muchas poblaciones similares a humanos las crearían;

(vi) Hay una buena probabilidad (del 25% o más) de que seamos simulaciones informáticas.²⁰

Pero ¿tiene sentido la hipótesis de la simulación? Si recordamos la épica película "Matrix", hay varias escenas en las que Neo, el personaje principal, se encuentra en una situación de total incertidumbre sobre si está en el mundo real o en el mundo simulado.

De hecho, si estamos en una simulación, será muy difícil, desde el punto de vista de nuestra subjetividad,

²⁰ Publicación original: Chalmers, D. (2022) *Reality+: Virtual Worlds and the Problems of Philosophy*, New York: W. W. Norton.

entender esto. Sin embargo, algunos pensadores argumentan que este argumento probabilístico falla porque algunas de sus premisas principales son poco plausibles.

Ahora veamos. Puede que algún día los ordenadores tengan conciencia, pero es improbable que su conciencia sea igual a la nuestra, dado que los mecanismos físicos de los ordenadores son muy diferentes de los mecanismos neurales, de carne y hueso, que producen la conciencia humana.

La suposición de que la conciencia artificial será igual que la conciencia humana presupone la independencia del sustrato: los estados mentales pueden operar en una amplia gama de sistemas físicos independientemente de su composición material.²¹

Sin embargo, podemos criticar esta suposición y defender que la conciencia depende efectivamente de un sustrato neurobiológico como el que la evolución "desarrolló" a través de nuestros cuerpos y cerebros (como veremos en varios diálogos en este libro). Y si ese es el caso, ningún sistema artificial podrá realmente reproducir la conciencia, al menos entendida como conciencia subjetiva y humana.

²¹ Sección basada en: Gouveia, S. & Neiva, D. (2017) "The Problem of Consciousness on the Mind Uploading Hypothesis" In *Philosophy of Mind: Contemporary Perspectives*, Cambridge: Cambridge Scholars Publishing.

Más aún, también podemos cuestionar la aceptación de que la segunda premisa (ii) sea falsa: de hecho, no es tan obvio por qué una futura civilización humana consideraría gastar innumerables recursos en simular miles de universos en lugar de aplicarlos en otras prioridades existenciales.

Además, esa misma civilización podría considerar que tales simulaciones son éticamente inmorales: si simular el universo implica simular seres humanos, y si la tesis de la independencia del sustrato es correcta, algo que actualmente no sabemos si es el caso, entonces simular a estos seres humanos también significaría simular su conciencia y, en consecuencia, su capacidad para sentir dolor y sufrir.

Si podemos simular esta capacidad sintiente, entonces podríamos estar cometiendo el mayor error ético en toda la historia de la humanidad (y sabemos que la lista de estos errores ya es bastante larga): podríamos estar simulando un número infinito de sufrimiento humano del cual tendríamos la obligación moral de evitar a toda costa.

Ya sea que consideres que el argumento a favor o en contra de la tesis de que estamos viviendo en una simulación es correcto, depende del lector. Pero recuerda tomar la decisión más sabia cuando te enfrentes a la elección entre la pastilla roja, que representa el mundo real, o la pastilla azul, que representa el mundo simulado.

Explorando el tema de las pastillas, a continuación, vamos a investigar, en compañía de la psicóloga y profesora Susan Blackmore, formas alternativas de pensar e investigar el fenómeno de la conciencia a través de varios estados alterados de conciencia y lo que pueden enseñarnos sobre la naturaleza de la mente consciente.

III. LA CONCIENCIA ALTERADA

El segundo diálogo cuenta con la participación de Susan Blackmore, escritora, conferenciante y profesora visitante en la Universidad de Plymouth, en el Reino Unido. Con un doctorado en Psicología, ella busca estudiar la naturaleza de la conciencia a través de metodologías más innovadoras.

Normalmente, al hablar de “conciencia”, presuponemos pensarla desde un punto de vista de su normalización, es decir, pensamos este fenómeno mental en su estado “normalizado”, cuando estamos despiertos experimentando las sensaciones del mundo, en lugar de estar soñando.

A lo largo de la historia registrada de las actividades humanas, se ha reconocido que existe una categoría de estados conscientes distintos de los estados normales a los que estamos acostumbrados. Estos estados pueden ser inducidos por el consumo de sustancias psicoactivas, la meditación o incluso, como exploraremos más adelante, los sueños lúcidos.

¿Por qué centrarse en estos estados que difieren de nuestras experiencias habituales? Porque pueden ser instrumentales para identificar y comparar información diversa cuando la conciencia está en un estado

alterado, lo que llamamos un “Estado Alterado de Conciencia” (EAC).

¿Cómo podemos intentar definir estos estados? Un método objetivo es especificar cómo se indujo ese estado. Por ejemplo, un cambio en la conciencia causado por el consumo de Ayahuasca conduce a un estado de conciencia diferente que un cambio inducido por la hipnosis o la meditación.



Imagen 5: “Essere” © Gaia Amoroso

Aunque esta es una propuesta interesante, se queda corta porque un estado inducido en el lector por la Ayahuasca, por ejemplo, tendrá efectos diferentes en su conciencia que si yo, el autor de este libro, consumiera la misma droga. No hay una relación causal y necesaria entre una sustancia y un EAC específico.

Otra forma para definir los EAC se basa en mediciones fisiológicas observadas que ocurren cuando se está bajo la influencia de sustancias, como la temperatura corporal, la frecuencia cardíaca, el consumo de oxígeno cortical, entre otros. Sin embargo, el desafío aquí radica en la variabilidad de estos criterios fisiológicos, que no permanecen consistentes entre todos los individuos, tanto en términos de experiencia subjetiva como en los métodos a través de los cuales se inducen.

Sabemos que cambios menores en la fisiología están asociados con fluctuaciones significativas en el estado subjetivo, y viceversa, por lo que no parece posible un mapeo directo. Por lo tanto, definir estos estados altamente volátiles, caracterizados por un inmenso impacto experiencial subjetivo, requiere un enfoque cuidadoso cuando se basa en elementos objetivos como la fisiología humana.

Hay algunas definiciones que podemos encontrar en manuales de psicología, neurociencia o filosofía como las siguientes:

- “(...) una alteración cualitativa en los patrones generales de funcionamiento mental de modo que el experimentador siente que sus operaciones de conciencia son radicalmente diferentes del funcionamiento ordinario”;²²
- “(...) un cambio temporal en el patrón general de experiencia subjetiva, de modo que el individuo cree que su funcionamiento mental es notablemente diferente de ciertas normas generales para su estado de vigilia normal”;²³
- “(...) existe un EAC siempre que haya un cambio de un patrón normal de funcionamiento mental a un estado que parece diferente para la persona que experimenta el cambio”.²⁴

Aunque estas definiciones son interesantes, todas ellas fallan en un aspecto importante: intentan definir los EACs comparándolos negativamente con los estados de conciencia "normales". El problema aquí es que también no tenemos una definición útil de estados de conciencia normalizados, lo que hace que este tipo de definiciones no nos ayuden a especificar

²² Publicación original: Tart, C. (1972) "States of consciousness and state-specific sciences", *Science*, 176: 1203-1210.

²³ Publicación original: Farthing, G. (1992), *The Psychology of Consciousness*, Englewood Cliffs, NJ: Prentice Hall.

²⁴ Publicación original: Nolen-Hoeksema, et al. (2014) *Atkinson & Hilgard's Introduction to Psychology*, Andover: Cengage Learning.

concretamente lo que podría ser considerada la naturaleza común de un estado alterado de conciencia.

Otra alternativa es explorar ejemplos concretos que puedan ayudar a aclarar esta dificultad conceptual. Comencemos, entonces, con aquellos estados que el lector más abierto de mente puede haber experimentado: los estados inducidos por drogas.

Las drogas psicoactivas son sustancias que afectan el funcionamiento mental o la conciencia. Operan modificando la acción de neurotransmisores endógenos o neuromoduladores.

Por ejemplo, estas drogas pueden afectar o aumentar a un neurotransmisor estimulando su liberación o bloqueando su recaptación para prolongar sus efectos. Por el contrario, pueden disminuir los efectos por inhibición o bloqueando su recepción en la membrana postsináptica.

Una razón por la que el impacto de las drogas que alteran la mente puede ser extenso es que un solo neurotransmisor puede estar activo en numerosas regiones del cerebro e incluso en todo el cuerpo.

Anfetamina es el estimulante más conocido y estudiado y tiene tres efectos principales en el cerebro:

- induce la liberación de serotonina;
- induce la liberación de dopamina;
- inhibe la recaptación de serotonina.

La serotonina juega un papel crucial en la regulación del estado de ánimo y el sueño, mientras que la dopamina ayuda a mediar el comportamiento motivado por la recompensa y las respuestas interpretativas hacia uno mismo, los demás y el entorno.

El consumo de anfetaminas incluye consecuencias como un aumento de la energía, aumento de sensaciones táctiles y aumento -tanto en frecuencia como en intensidad- de sentimientos de amor y empatía hacia los demás.²⁵ La experiencia y sus efectos a menudo dependen del entorno en el que se consume la droga. El uso frecuente puede llevar a la tolerancia y, consecuentemente, a la adicción a largo plazo.

Sin embargo, en un contexto terapéutico y de baja frecuencia, puede emplearse para tratar condiciones como el trastorno de estrés postraumático y la ansiedad social con cierta eficacia, en comparación con tratamientos más convencionales.²⁶

La droga psicodélica más común es el cannabis, que contiene aproximadamente 85 componentes cannabinoides. Describir los efectos subjetivos del

²⁵ Publicación original: Holland, J. (ed.) (2001) *Ecstasy: The complete guide: A comprehensive look at the risks and benefits of MDMA*, Rochester, VT: Park Street Press.

²⁶ Publicación original: Danforth, A., et al. (2016) *MDMA-assisted Therapy: a new treatment model for social anxiety in autistic adults*, *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 64: 237-249.

cannabis no es una tarea fácil, en parte porque varían enormemente de una persona a otra.

Sin embargo, la investigación científica ha revelado algunos efectos comunes, como respuestas emocionales, que incluyen euforia y relajación en dosis más bajas, y miedo y paranoia en dosis más altas.

Los efectos sensoriales incluyen una mayor percepción de profundidad, una sensación de perfeccionamiento de nuestros sentidos, un aumento de la capacidad de respuesta y del placer sexual, una desaceleración del tiempo y ampliación del espacio, además de una mayor concentración en el presente.

Algunas personas encuentran un sentido de lo sagrado o divino en el consumo de esta droga. Sabemos también que, en algunas personas, el consumo puede llevar a un aumento de la creatividad, mientras que en otras puede llevar a un pensamiento lento con efectos negativos en la memoria a corto plazo. En dosis altas, algunos individuos reportan fenómenos de sinestesia, donde se intercambian elementos sensoriales, como por ejemplo ver un color al escuchar un sonido.

Ayahuasca es otro tipo de psicodélico que se está volviendo cada vez más común y popular. Considerada una sustancia curativa, inicialmente induce episodios de vómitos, y después de unos minutos, surge una variedad desconcertante de sensaciones corporales,

transformaciones, visiones y percepciones.²⁷La contemplación de la muerte es común, así como las ideas místicas sobre asuntos personales y preguntas existenciales profundas.

Durante el consumo de esta sustancia, el cerebro experimenta alteraciones significativas, como una disminución en la conectividad entre el parahipocampo y el córtex retrosplenial, que se correlaciona con relatos de disociación del “yo” o que explica por qué quienes pasan por una experiencia con ayahuasca relatan, muchas veces, haber sentido que su “yo” se ha diluido o fusionado con el ambiente en el que se encontraban.²⁸

Finalmente uno de los más potentes alucinógenos es el LCD (dietilamida de ácido lisérgico). Los efectos de esta sustancia abarca una amplia gama de experiencias que incluyen sensaciones positivas como alegría y euforia, así como sensaciones negativas como terror, desesperación y desintegración del “yo”.

En los primeros estudios de imagen cerebral controlado por placebo, los participantes recibieron 75

²⁷ Publicación original: Luna, L. e White, S. (2016) *Ayahuasca Reader: Encounters with the Amazon's Sacred Vine*, Santa Fe: Synergetic Press.

²⁸ Original publication: Uthaug, M. et al. (2018) “Sub-acute and long-term effects of ayahuasca on affect and cognitive thinking style and their association with ego dissolution”, *Psychopharmacology*, 235: 2979-2989.

microgramos de LSD por vía intravenosa. Fue detectado un aumento de conectividad funcional en todo el cerebro, mientras los efectos locales coincidieron con cambios en la experiencia. Por ejemplo, alucinaciones visuales fueron correlacionadas positivamente con un aumento de flujo sanguíneo cerebral y de conectividad funcional en el córtex visual primario (V1).

Todas esas sustancias traen algo nuevo para aquellos que la consumen: una experiencia consciente diferente y anormal que puede ayudar a entender la naturaleza de la conciencia. ¿Porqué? Porque nosotros podemos comparar, por ejemplo, un cerebro normal de un sujeto sin inducción de droga con un cerebro inducido por droga, lo que nos permitirá obtener un conocimiento relevante acerca de la actividad neuronal, así como del dinamismo de las redes neuronales o saber más sobre la relación de determinadas zonas del cerebro.



Imagen 7: "Siamo uguali" © Gaia Amoroso

Ahora veamos una nueva forma de pensar los "estados alterados de conciencia", en esta ocasión a través de la meditación. ¿Será que una meditación puede ser considerada como un EAC? Algunas definiciones parecen implicar que si:

- "(...)la meditación puede ser considerada como un proceso lento, acumulativo y de largo plazo

para producir un estado alterado de conciencia".²⁹

Algunos practicantes del budismo pueden tener experiencias de kenshō (despertar), incluyendo vislumbres dentro de la supuesta naturaleza de la mente. También sabemos que quien medite siente que su mente ha sido radicalmente alterada. Y en este sentido esto puede ser considerado EAC.

Sin embargo, algunos científicos argumentan que la meditación no es nada más que una forma particular de... dormir. Sin embargo, el neuropsiquiatra Peter Fenwick³⁰ mostró que los perfiles EEG en la meditación no son exactamente similares a aquellos de sueño o somnolencia, aunque muchos practicantes entran en un estado de "microsueño" durante la meditación. En otro estudio, los practicantes durmieron alrededor de un tercio del tiempo durante la meditación.³¹

Los efectos benéficos de la meditación, como reducir la ansiedad, la depresión y mejorar el rendimiento cognitivo, pueden ser explicados en parte por la semejanza con los beneficios conocidos de los

²⁹ Publicación original: Wallace, B. e Fisher, L. (1991) *Consciousness and Behavior*, Boston, MA: Allyn and Bacon.

³⁰ Publicación original: Fenwick, P. (1987) "Meditation and the EEG" In M. West (ed.) *The Psychology of Meditation*, Oxford: Clarendon Press.

³¹ Publicación original: Austin, J. (1998) *Zen and the brain: Toward an Understanding of Meditation and Consciousness*, Cambridge, MA: MIT Press.

microsueños. No obstante, y de manera interesante, muchos practicantes reportan ser capaces de distinguir entre la meditación profunda y los estados de sueño, aunque esta distinción es difícil de explicar objetivamente.

Hablando de sueño, otra manera de considerar los “estados alterados de conciencia” es a través de los sueños lúcidos. Un sueño lúcido es caracterizado por la percepción de que tú estás soñando en ese preciso momento. Todos los días cuando dormimos, transitamos por un ciclo de tres estados: vigilia, REM (rapid eye movement) y sueño no REM. El sueño típico de una noche consta de cuatro o cinco ciclos entre el sueño no REM y el sueño REM, a menudo incluyendo algunos microdespertares inconscientes

Los estados de vigilia y sueño son caracterizados por indicadores del comportamiento, como la rapidez de despertar, los movimientos oculares, la tensión muscular y la actividad cerebral, que pueden medirse utilizando métodos como el electroencefalograma (EEG), entre otros.

Sabemos que en el sueño REM, el cerebro está bastante activo y su EEG se asemeja al del cerebro en estado de vigilia. En el sueño no REM, la tasa global de disparo de las neuronas es tan alta como en los estados de vigilia, pero el patrón es bastante diferente, con el EEG dominado por ondas largas y lentas en lugar de complejas y rápidas.

Durante el sueño, el cerebro se aísla de varias maneras: hay una inhibición de la entrada sensorial a nivel tálamo-cortical durante el sueño no REM, mientras que en la fase REM, esta inhibición es más periférica. Estudios con EEG y fMRI muestran que durante estímulos auditivos en la fase REM, la corteza auditiva permanece activa, y en la fase REM intermitente, que incluye movimientos oculares y espasmos musculares, el cerebro opera en un circuito cerrado y se encuentra funcionalmente aislado del mundo exterior.³²

En el sueño REM, el tronco encefálico bloquea los comandos motores en la médula espinal, lo que evita que las actividades mentales se traduzcan en movimientos físicos. Sin embargo, si este mecanismo falla, esto podría explicar por qué el tío del lector suele levantarse en medio de la noche para comerse todos los postres del refrigerador. El sonambulismo puede atribuirse a una falla en este mecanismo.

En esta fase REM, la amígdala, el hipocampo y el cíngulo anterior muestran una actividad elevada, así como partes del sistema visual y áreas de asociación visual. La corteza prefrontal dorsolateral, asociada con funciones ejecutivas como la memoria de trabajo, la resolución de problemas y la organización motora,

³² Publicación original: Wehrle, R. et al. (2007) "Functional microstates within human REM sleep: First evidence from fMRI of a thalamocortical network specific for phasic REM periods", *European Journal of Neuroscience*, 25: 863-871.

muestra una actividad ,ás reducida en comparación con el estado de vigilia.

La investigación sobre los sueños lúcidos proporciona un contexto interesante para explorar los correlatos neuronales de la conciencia, ya que nos permite correlacionar variables fisiológicas, neuroquímicas y comportamientos con descripciones subjetivas de los sueños, y puede ayudarnos a tener una mejor comprensión de las dimensiones de la experiencia consciente.

Por último, una forma final de pensar sobre la naturaleza de la conciencia es a través de algo muy especial que puede suceder en nuestra mente en un momento dado, como le ocurrió a la profesora Susan Blackmore, quien, a los 19 años, se encontró experimentando una “experiencia fuera del cuerpo” (EFC) (cf. Diálogo con Susan Blackmore).

Una EFC es una experiencia en la que una persona parece percibir el mundo desde una ubicación fuera de su cuerpo físico. Durante una EFC, los individuos pueden observarse a sí mismos desde una perspectiva externa a su cuerpo.



Imagen 8: "Esperienza" © Gaia Amoroso

Aquellos que tienen EFC a menudo reportan más experiencias psíquicas y una mayor creencia en lo paranormal en comparación con aquellos que no las tienen. Además, los individuos con EFC pueden tener una mejor capacidad para recordar sueños y una mayor frecuencia de sueños lúcidos.³³

¿Qué puede decirnos la EFC sobre la conciencia? Algunas personas interpretan estas experiencias inusuales como una prueba absoluta de que la

³³ Publicación original: Blackmore, S. (2017) *Seeing Myself: The New Science of Out-of-body Experiences*, London: Robinson.

conciencia es un alma inmaterial, independiente del cuerpo o del cerebro. Aunque es comprensible que alguien con muy poco conocimiento científico pueda creer en esa tesis, hay otras explicaciones posibles.

Veamos, si el alma realmente pudiera ver el mundo físico durante la EFC, esto implicaría una interacción con el mundo físico, contradiciendo su supuesta naturaleza no física y, por lo tanto, debe ser una entidad física detectable. Sin embargo, esto contradice la supuesta naturaleza no física del alma. Por otro lado, si el alma es no física, entonces no puede interactuar con el mundo físico para poder observar el cuerpo (físico) desde un punto de vista externo.

De hecho, una perspectiva alternativa a las teorías dualistas sugiere que, a pesar de la experiencia subjetiva de salir del cuerpo durante una EFC, en realidad nada sale de nuestro cuerpo físico.

Por ejemplo, algunas de las teorías psicoanalíticas han propuesto que las EFC podrían reflejar el miedo a la muerte, la regresión del ego o la re-experimentación del trauma del nacimiento. Sin embargo, estas teorías psicoanalíticas son difíciles de probar científicamente y han tenido un impacto limitado en el avance de nuestra comprensión del fenómeno de las EFC.

Sin embargo, si hablamos de la neurociencia más actual, descubrimos que el lóbulo temporal desempeña un papel significativo en las EFC, ya que, en pacientes

epilépticos, la estimulación del lóbulo temporal provoca episodios de este tipo, así como experiencias psíquicas y místicas.

El neurocientífico Michael Persinger³⁴ propuso que las creencias y experiencias místicas eran "creaciones" de la función del lóbulo temporal de nuestro cerebro, habiendo logrado inducir varias EFC y distorsiones corporales al utilizar la Estimulación Magnética Transcraneal que, al centrarse en el lóbulo derecho, produjo auténticas EFC en los pacientes.

El área específica del cerebro involucrada en la estimulación es la unión temporoparietal derecha, lo que hace sentido para tener un impacto en las experiencias extracorporales (EFCs), dado que esta área es responsable de procesar la información visual, táctil, propioceptiva y vestibular. Estos insumos, cuando se combinan, juntas forman el esquema corporal para cada uno de nosotros.

De esta manera, podemos ofrecer una explicación científica para un fenómeno que puede parecer místico, pero que tiene, en su base, una naturaleza perfectamente humana.

A continuación, presentaremos otras ideas de la psicología informadas por la neurociencia, esta vez

³⁴ Publicación original: Persinger, M. (1983) *Religious and mystical experiences as artifacts of temporal lobe function: A general hypothesis*, *Perceptual and Motor Skills*, 57: 1255-1262.

desde una psicología de influencias más evolutiva, para intentar explicar por qué nosotros, como seres humanos y algunos animales, tenemos experiencias subjetivas conscientes.

IV. CONCIENCIA EVOLUTIVA

El tercer diálogo presenta una contribución del renombrado psicólogo y teórico Nicholas Humphrey, profesor de Psicología en la London School of Economics. En esta discusión, encontrarás ideas fascinantes sobre la naturaleza de la conciencia desde una perspectiva evolutiva.

La historia de Nicholas Humphrey en el estudio de la conciencia comenzó, como verás en el diálogo, cuando él tenía solo 23 años y era estudiante de doctorado en un laboratorio de psicología en la Universidad de Cambridge.

Durante su investigación, inicialmente investigó a un mono que, mientras estaba anestesiado, tenía un electrodo insertado en su cerebro, específicamente en el colículo superior, un área "antigua" en la evolución neuroanatómica del cerebro responsable del procesamiento visual.

El colículo superior precede a la corteza visual más desarrollada que permite la percepción visual consciente en humanos y mamíferos. El aspecto curioso de este episodio es que, aunque el mono no estaba despierto, las células nerviosas en el colículo superior estaban activas, sugiriendo que, tal vez, el

procesamiento visual estaba ocurriendo sin una sensación consciente asociada.

En algún momento posterior, Humphrey se encontró con un mono llamado Helen, cuya corteza visual había sido completamente removida, dejando solo el colículo superior. A través de varias interacciones durante muchos meses, Helen, quien debería haber estado completamente ciega, aparentemente desarrolló la capacidad de ver nuevamente. Podía tomar su fruta favorita, reconocer al profesor y realizar otras hazañas notables que no deberían haber sido posibles para un mamífero sin visión.

Ahora, el autor de "Soul Dust" comenzó a sospechar que Helen podría estar teniendo percepciones visuales sin tener experiencias conscientes asociadas, podría procesar información de su entorno sin formar una imagen consciente de esa información.

Toda esta investigación llevó a su asesor doctoral, el profesor Larry Weiskrantz, a observar a un paciente humano cuya corteza visual se redujo a la mitad debido a un accidente. Este paciente podía identificar objetos con alta precisión que estarían presentes en su campo visual, aunque el paciente admitió no haber visto nada por sí mismo.



Imagen 9: "Everything is worth it" © Gaia Amoroso

El famoso fenómeno de la "visión ciega" fue descubierto. Este fenómeno nos ayudó a comprender mucho sobre el funcionamiento de la visión, pero también planteó preguntas fundamentales sobre la naturaleza de la conciencia. Todo este episodio despertó un enorme estímulo y curiosidad en el profesor Humphrey, motivándolo a contribuir a aclarar la existencia de experiencias conscientes en el mundo.

La principal contribución de Humphrey es avanzar en una teoría de la conciencia que pueda explicarse a través de la teoría evolutiva darwiniana. En palabras del autor:

“(…) Argumentaré que la verdad sobre la conciencia, si y cuando la veamos desde la perspectiva correcta, es que de hecho es el producto de una ingeniería biológica altamente improbable: una maravillosa obra de arte de la naturaleza que da lugar a todo tipo de impresiones misteriosas en nuestras mentes, pero algo que tiene una explicación física relativamente sencilla.”³⁵

Y así comienza esta maravillosa historia de cómo la conciencia podría haber evolucionado a través de la selección natural.

El primer paso es argumentar que la conciencia tiene un impacto en nuestro comportamiento. ¿Y por qué el comportamiento? Dado que la conciencia, según la entendemos, es una característica inherente a la vida en la tierra, podemos suponer que, al igual que cualquier otra característica especializada de los organismos vivos, evolucionó porque confiere alguna ventaja selectiva.

De una forma u otra, la conciencia debe ayudar al organismo a sobrevivir y reproducirse. Esto solo puede ocurrir si, de alguna manera, está alterando la forma en que el organismo interactúa con el mundo exterior, un proceso que típicamente implica comportamiento.

³⁵ Original publication: Humphrey, N. (2011) *Soul Dust: the Magic of Consciousness*, Princeton: Princeton University Press.

Pero, ¿De qué forma la conciencia distingue a un organismo en términos de comportamiento? Cuando examinamos "criaturas conscientes", como las denomina Humphrey, nos damos cuenta de que ellas no poseen características físicas distintivas en comparación con criaturas no conscientes. La conciencia no proporciona una mayor salud, fuerza o belleza; más bien, parece ejercer sus efectos en la sobrevivencia de un organismo a través de lo que podemos llamar la "psicología" de la criatura.

Por lo tanto, ser fenomenalmente consciente debe estar relacionado con la forma en que un organismo piensa, desea o cree, y el hecho de ser consciente lo lleva a actuar de una forma adaptativa al mundo a través de comportamientos específicos que confieren una ventaja y que pueden ser identificados.

Sin embargo, a pesar de esta "ventaja" adaptativa, aún no está claro por qué razón la conciencia fue seleccionada como un elemento relevante por la selección natural. Este es el punto central de la teoría evolutiva de la conciencia: si la selección natural puede "ver" los efectos, cualesquiera que sean, del cambio psicológico en el comportamiento, presumiblemente otros observadores externos también pueden percibir las ventajas adaptativas de ser consciente.

Pero, ¿cómo la conciencia emergió evolutivamente para convertirse en una característica fundamental de los seres humanos hoy en día? Humphrey cree que la

aparición de la conciencia ocurrió rápidamente y fue un fenómeno de "todo o nada". Apareció más tarde en la historia evolutiva cuando nuestros antepasados homínidos desarrollaron diversas habilidades sociales como la imitación, el engaño y el lenguaje.

Así, podemos argumentar que la conciencia es una propiedad emergente que evolucionó por su función social, por la habilidad de entender, predecir y manipular el comportamiento de otros individuos. La evolución "favoreció" a los individuos que poseían estas capacidades sobre aquellos que no las tenían, llevando al desarrollo y la persistencia de la conciencia.

Al igual que las especies de grandes simios que viven hoy en día, los humanos siempre han vivido en grupos sociales complejos. Conocer las intenciones de otros individuos puede ser extremadamente útil en determinar quién ocupa una posición más alta en la jerarquía social, en quién podemos confiar, así como saber con quién podemos formar alianzas, entre otras consideraciones sociales.

Así, de acuerdo con Humphrey, los antepasados que podían entender, predecir y manipular el comportamiento de otros tenían una clara ventaja adaptativa. En este punto, se podría argumentar que los humanos podrían haber adquirido esas habilidades por simplemente observar el comportamiento de los demás y sus consecuencias desde un punto de vista externo, algo similar a los conductistas.

Pero el profesor emérito de la Universidad de Cambridge cree que podría haber una mejor manera de lograr este resultado. A través de la hipótesis de que los individuos adquirieron la capacidad de introspección, de ponerse en el lugar de otra persona e intentar "observar" sus propias mentes.

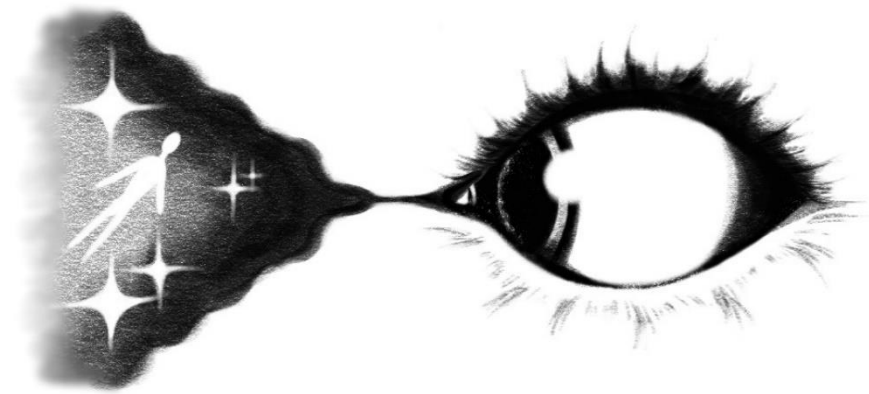


Imagen 10: "Percezione" © Gaia Amoroso

Humphrey compara esta capacidad con un "ojo interno", direccionado no por el mundo exterior, como la mayoría de los órganos sensoriales, sino más bien direccionado hacia al mundo interior del individuo. Este "ojo interno" no puede observar el funcionamiento neuronal del cerebro, sino que en vez de eso percibe

una versión psicológica más accesible de esa actividad, los estados conscientes subjetivos.

¿Podría esto significar que la conciencia puede considerarse una "invención", en el sentido de que no existía en un cierto período del desarrollo humano, y que surgió más tarde?

Para Humphrey, la respuesta es doble: la conciencia es una facultad cognitiva desarrollada por la selección natural, diseñada para ayudarnos a hacer sentido a nosotros mismos y a nuestro entorno. Simultáneamente, es una fantasía creada por el cerebro, diseñada para alterar el valor que atribuimos a nuestra existencia.

Sin embargo, si este es el caso, corremos el riesgo de cometer el error del "teatro cartesiano", como señala Daniel Dennett. Esta idea sugiere que hay un lugar específico en el cerebro donde se "filma" lo que acontece en el mundo exterior.

Este enfoque de la conciencia no implica necesariamente una tesis ilusionista de la conciencia - la idea de que la conciencia es una ilusión creada por el cerebro- esto es porque lo que acontece dentro del "teatro" no es una réplica del mundo exterior, sino más bien del mundo "interior", esto es, el mundo de la subjetividad y los "qualia".

Es cierto que el autor de "Sentience" comienza por negar el "realismo" de los "qualia", tal como fueron

definidos por filósofos como David Chalmers o Thomas Nagel, quienes, como vimos anteriormente, ellos consideran los "qualia" como existentes de forma fundamental e independiente, acercándose a una teoría ilusionista.

Este enfoque argumenta que, aunque las experiencias subjetivas conscientes parecen poseer estas maravillosas propiedades no físicas, esto solo puede ser así porque el cerebro nos está jugando una mala pasada.

Como enfatiza Daniel Dennett, considerado uno de los principales defensores de la teoría ilusionista: "La conciencia es una ilusión del cerebro, para el cerebro, por el cerebro".³⁶

Entonces, cuando tenemos una experiencia subjetiva consciente, nuestro propio cerebro está "montando" un espectáculo de magia que nos hace creer que estamos experimentando algo fuera de nosotros mismos. De cierta manera, este enfoque socava el misterio de la experiencia humana, la noción de que somos únicos, mágicos y excepcionales en el universo.³⁷

³⁶ Publicación original: Dennett, D. (1991) *Consciousness Explained*, New York: Little Brown.

³⁷ Publicación original: Humphrey, N. (2020) "The Invention of Consciousness", *Topoi*, 39: 13-21.

Pero para el autor de "Soul Dust," esto también puede ser una ventaja, tan pronto como nos demos cuenta que no hay nada especialmente misterioso sobre la existencia de la conciencia, podemos utilizar esta información como un recurso para "enriquecernos" a nosotros mismo como seres humanos. Este enriquecimiento puede ocurrir a través de las artes, a través de la ciencia y trascendiendo una existencia meramente "terrenal".

Para concluir esta nota introductoria, echemos un vistazo a las últimas palabras del capítulo 23 del nuevo libro de Humphrey titulado "Sentience", centrado en la excepcionalidad de la conciencia en el mundo:

"Aunque la idea de que una característica evolucionada naturalmente con "intención" esté equivocada, me imagino que Darwin mismo podría haber visto la conciencia fenomenal como un logro "último", la gloria máxima del proceso evolutivo que comenzó con el Big Bang. Es una invención tan sublime que, si dejara de existir, ciertamente disminuiría la totalidad de la creación.."

Esta visión evolutiva de la conciencia ofrece un enfoque interesante que puede ser estudiada empíricamente. Sin embargo, no aclara específicamente cómo nuestro

cerebro produce la propiedad emergente de la subjetividad consciente.³⁸

Para cerrar esta primera parte del libro dedicada a la conciencia, presentaré las ideas del último invitado con quien debatí sobre este tema, el profesor Sir Roger Penrose.

³⁸ Sección basada en: Humphrey, N. (2023) *Sentience: the Invention of Consciousness*, Cambridge, MA: MIT Press.

V. LA CONCIENCIA CUÁNTICA

El cuarto y último diálogo es, quizás, aquel que para el lector podría tener mayor dificultades de seguir. De tal modo que para prepararlo para una de las conversaciones más estimulantes en este libro, intentaré introducir algunos de los conceptos centrales desarrollados por el premio nobel de Física, Sir Roger Penrose, profesor emérito en la Universidad de Oxford.

¿Cómo podemos pensar en la conciencia? Algunos científicos y filósofos consideran la conciencia como una propiedad emergente de la computación entre neuronas que se interconectan y cambian en sinapsis mediadas químicamente. El problema de la conciencia, en este contexto, es el desafío de comprender las computaciones específicas que ocurren y si podemos replicar ese tipo de procedimiento.

Desde esta perspectiva, el cerebro se ve como un tipo de computadora digital que procesa información del entorno hacia la mente, y la conciencia se piensa que desempeña algún tipo de papel crucial en este proceso.

Sin embargo, la noción de "computación" en este sentido clásico es bastante problemática; para el físico británico, el cerebro opera en una escala de

procesamiento completamente diferente a la de una computadora digital actual.

Por lo tanto, el enfoque de Penrose presupone ampliar el alcance del poder explicativo desde la física clásica/neurociencia computacional tradicional (es decir, la neurona) hacia un nivel "más pequeño".

A través de ciertos elementos de la física cuántica, se propone una teoría de la conciencia con una explicación cuántica. Para alcanzar este paso, Penrose critica la idea de que las propiedades cuánticas dejan de ser relevantes cuando "subimos" la escala de complejidad en la biología. Este argumento sugiere que la neurona, las interacciones neuronales, el cerebro, etc., no son sujetos apropiados para ser interpretados por la mecánica cuántica (MC), siendo que deben ser interpretados por la física clásica.

Contrario a esta tesis un grupo de Científicos³⁹ mostró que MC desempeña un papel fundamental en la fotosíntesis, descubriendo que en las células de clorofila, la transferencia de luz alcanza una eficiencia cercana al 100 por ciento, lo que supera con creces cualquier tipo de tecnología moderna humana.

³⁹ Publicación original: Engel, G. et al. (2007) "Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems", *Nature*, 446 (2007): 782-786.

Además, también sugiere que debe haber algún tipo de "espacio seguro" de MC para mantener la coherencia del sistema y lograr este tipo de eficiencia.

Penrose argumenta que necesitamos ir más allá de la computación digital para alcanzar una teoría viable de la conciencia porque la física clásica tiene tres suposiciones que impiden una explicación de la conciencia humana:

- Determinismo causal: podemos conocer todas las condiciones iniciales de los sistemas físicos y, por lo tanto, podemos predecir su comportamiento futuro basándonos en una cadena causal;
- Localidad/independencia: dos sistemas separados en el espacio no pueden interactuar instantáneamente;
- Realismo objetivista: los objetos en la realidad externa existen con propiedades bien definidas que son independientes de cualquier observador.

Sin embargo, el Premio Nobel en Física argumenta que la conciencia es no computacional (es decir, no algorítmica) como consecuencia directa del teorema de Gödel.⁴⁰

⁴⁰ Publicación original: Gödel, K. (1931) "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I", *Monatshefte für Mathematik und Physik*, 37: 349–360.

Penrose demostró que la cualidad mental de "comprender" (*understanding*) no puede ser encapsulada por ningún sistema computacional y debe derivar de algún efecto "no computable". Sugiere que el ingrediente no computable necesario para la conciencia humana y la comprensión humana debe encontrarse en un área donde nuestras teorías físicas actuales están fundamentalmente incompletas, aunque de importancia relevante para las escalas pertinentes al funcionamiento de nuestros cerebros.

La única posibilidad "seria" es la incompletitud de la Teoría Cuántica, una incompletitud que tanto Albert Einstein como Erwin Schrödinger ya habían reconocido, referida como el Problema de la Medición. Una forma de resolver este problema sería proporcionar una extensión de la estructura estándar de la MC, introduciendo una forma objetiva de reducción del estado cuántico, denominada "reducción objetiva" (RO).

¿Significa esto que la conciencia reduce los estados cuánticos? En esta versión, que llamamos "la interpretación clásica" de la MC, es el acto de observación por parte de un sujeto lo que define el estado cuántico y viola el principio de superposición: el gato de Schrödinger estaría muerto o vivo hasta que alguien (por ejemplo, un observador) observe el

sistema, definiendo la superposición en una posición particular.⁴¹

Para entender la superposición desde la perspectiva clásica, imagina lanzar una moneda para decidir qué vas a cenar, si cae en cara, será seitán empanizado; si cae en cruz, será raclette. Justo después de lanzar la moneda al aire, durante los pocos segundos que está en el aire, la moneda estará en un estado de superposición, es decir, está simultáneamente en ambas posiciones, cara y cruz.

La cara de la moneda solo se define cuando el lector recoge la moneda y observa el lado que cayó (¡que es cara!). Las partículas subatómicas son como esta moneda, existen en estados de superposición hasta que una observación hace que colapsen en un estado determinado.

⁴¹ Publicación original: Schrödinger, E. (1935) "Die gegenwärtigen situation in der quantenmechanik", *Naturwissenschaften*, 23: 807-812, 823-8, 844-9.

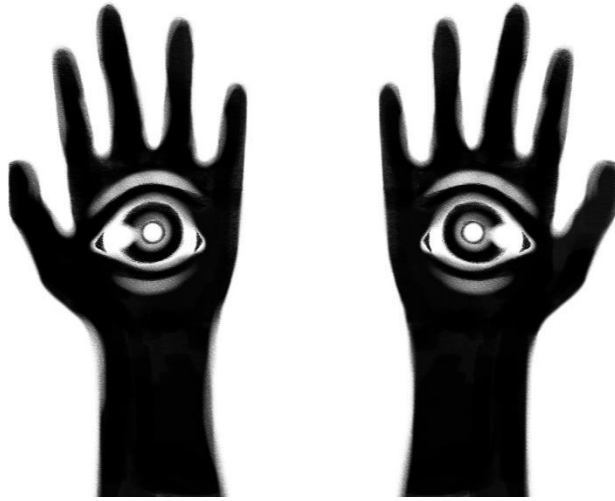


Imagen 11: "Riflessione" © Gaia Amoroso

Sin embargo, la visión de Penrose es precisamente lo opuesto a esta descripción. Para él, existe un conflicto entre la mecánica cuántica y la teoría de la relatividad general de Einstein que pide que algo "nuevo" suceda.

42

En la opinión del premio nobel de Física, la reducción del estado no ocurre por observación, sino espontáneamente, y es en este proceso donde aparece la conciencia. Pero, ¿tiene sentido esto? Notece que, el nivel de la MC solo funciona como teoría a un nivel muy "pequeño" (subatómico), y el nivel de procesamiento de

⁴² Publicación original: Einstein, A. (1916) "Die Grundlage der allgemeinen Relativitätstheorie", *Annalen der Physik*, 49 (7): 769-822.

información neuronal ocurre a una escala mucho mayor.

Si la física cuántica se extiende a niveles "elevados", se encontraría en un contexto "caliente", "húmedo" y "ruidoso" debido a un contacto con el ambiente, lo que necesariamente lleva al colapso de los estados cuánticos y a su decoherencia cuántica.

Esta decoherencia se puede explicar con el siguiente ejemplo: imagina que el gato del lector es un gato cuántico que está simultáneamente vivo y muerto según el principio de superposición. Si tu gato interactúa con partículas del entorno (por ejemplo, con fotones), ocurrirá la decoherencia. Dado que el entorno lo "observa" indirectamente, esta observación haría que tu gato asumiera un estado definido, vivo o muerto, en lugar de existir en ambos estados al mismo tiempo (con suerte, por supuesto, es la primera opción cuántica).

Por lo tanto, siguiendo esta línea de razonamiento, parece improbable que los principios de la física cuántica operen al nivel del intercambio de información entre las neuronas en el cerebro y produzcan conciencia. ¿Representa esto un obstáculo para la teoría de la conciencia de Penrose?

De acuerdo con el físico de Oxford, esta crítica no se sostiene debido a intrigantes razones evolutivas, a lo largo de millones de años, la biología ha evolucionado

para abordar el desafío de la decoherencia cuántica. Esta adaptación se logra mediante el desarrollo de una estructura subatómica dentro de las neuronas llamada "microtúbulos". Su estructura única garantiza la coherencia cuántica, similar al proceso natural observado en la fotosíntesis, como se mencionó anteriormente.

La referencia a los microtúbulos no es original de Penrose, sino que proviene de su colega, el anesthesiólogo estadounidense Stuart Hameroff. Hameroff propuso a Penrose que la coherencia cuántica en el cerebro podría ocurrir dentro de estos microtúbulos. Los microtúbulos son estructuras proteicas con una formación tubular dentro de las células eucariotas (parte del citoesqueleto). Desempeñan varios roles, como determinar la forma de la célula, coordinar el movimiento y supervisar la división celular.

Hameroff sugiere que los microtúbulos son el dispositivo cuántico que Penrose buscaba en su teoría, ya que ayudan a controlar la fuerza de las conexiones sinápticas, y su forma tubular puede protegerlos del ruido circundante de la neurona más grande. Imagina que las neuronas son pequeñas fábricas: los microtúbulos serían las vías que ayudan a guiar y organizar el movimiento dentro de estas fábricas.

Sin embargo, se necesita más que una simple serie continua de momentos aleatorios de coherencia

cuántica para tener algún impacto en la conciencia: son estos momentos de conciencia, orquestados por los microtúbulos en nuestros cerebros, los que tienen la capacidad de almacenar y procesar información y memoria.

Para Hameroff y Penrose, los microtúbulos pueden preservar adecuadamente la coherencia cuántica hasta que alcancen el nivel neuronal: para la conciencia, es necesario que muchos microtúbulos en varias neuronas diferentes actúen de manera "orquestada".

¿Cuál es, entonces, la diferencia entre los estados cuánticos normales y los estados cuánticos que llevan a la conciencia? Para este par de científicos, la clave radica en la coherencia global, por lo que abogan por la reducción objetiva (RO), en la cual los estados cuánticos colapsan en una opción u otra.

La parte "orquestada" se propone para asegurar que los esfuerzos colaborativos entre múltiples microtúbulos son necesarios para influir en el nivel neuronal.

Para resumir las ideas hasta ahora, tenemos la siguiente configuración teórico-conceptual:

- Se propone un modelo en el cual la conciencia emerge a través de efectos cuánticos que ocurren dentro de las estructuras subcelulares internas a las neuronas, conocidas como microtúbulos;

- Este modelo postula los llamados "colapsos objetivos" que implican que el sistema cuántico transite de una superposición de varios estados posibles a un estado único definido, pero sin la intervención de un observador o medición, como ocurre en la mayoría de los modelos de mecánica cuántica;
- Según Penrose y Hameroff, el entorno interno de los microtúbulos es particularmente adecuado para tales colapsos objetivos. Los auto-colapsos resultantes producen un flujo coherente que regula la actividad neuronal y posibilita procesos mentales no algorítmicos.

Después de todo esto, es importante ahora introducir la llamada "Interpretación de Penrose" de la mecánica cuántica, ya que es fundamental para comprender su teoría de la conciencia.

La interpretación de Penrose es una especulación sobre la relación entre la Mecánica Cuántica (MC) y la relatividad general de Einstein. Propone que un estado cuántico permanece en superposición hasta que la diferencia en la curvatura del espacio-tiempo alcanza un nivel significativo para que colapse, un fenómeno al que se refiere como "auto-colapso".

Esta perspectiva es una alternativa a la "Interpretación de Copenhague", que postula que la superposición falla cuando se realiza una observación (pero que no es

objetiva en su naturaleza), y también es una alternativa a la "Interpretación de los *Muchos Mundos*", que afirma que los resultados alternativos de una superposición son igualmente "reales", mientras que su decoherencia mutua impide interacciones observables subsecuentes.

La interpretación de Penrose rechaza estas dos interpretaciones dependientes del observador (ambas siendo teorías subjetivas), abogando por una forma de teoría de colapso objetivo. Para esta teoría, la función de onda se considera una onda física, y el colapso de la función de onda se postula como un proceso físico, sin que los observadores jueguen un papel especial o causal.

Penrose teoriza que la función de onda no puede mantenerse en una superposición más allá de una cierta diferencia de energía entre los estados cuánticos, y que este umbral estará relacionado a las influencias gravitacionales con las partículas.

La propuesta de "Reducción Objetiva" (RO) tendría su inicio determinado por una condición denominada el criterio "de un gravitón". La propuesta de Diósi-Penrose proporciona un umbral físico objetivo, indicando una vida plausible para los estados cuánticos superpuestos. Esta propuesta sugiere que cada evento de RO, que es un proceso puramente físico, es en sí mismo un tipo

primitivo de "observación", un momento de "experiencia protoconsciente."⁴³

Para lograr esto, es necesario que la superposición cuántica evite la decoherencia ambiental inmediata y persista hasta que se alcance un límite de tiempo específico. Esto se logra argumentando que una superposición cuántica es:

- "orquestada", es decir, adecuadamente organizada, imbuida de informaciones cognitivas y capaz de integración y computación;
- Aislada de un entorno aleatorio no-orquestado durante el tiempo suficiente para que la superposición evolucione hacia el colapso y, en ese momento, cree un instante de conciencia.

El problema con este enfoque es que se considera que la ecuación de Schrödinger describe el formalismo cuántico de un sistema a temperatura cero: sería absurdo considerar que un cerebro consciente está en un entorno térmico cercano a cero (¡por el contrario, está lejos de eso!).

También sabemos que las computadoras cuánticas actuales requieren temperaturas muy cercanas a cero

⁴³ Publicación original: Hameroff, S. & Penrose, R. (2014) "Consciousness in the universe: A review of the 'Orch OR' theory" *Physics of Life Review*, 11 (1): 39-78.

grados en la escala Kelvin para ser funcionales. Por lo tanto, se puede argumentar que, considerando las actividades cuánticas relevantes en el cerebro a temperaturas lejos de cero, esto puede ser altamente problemático y plantear un obstáculo para toda la teoría avanzada propuesta por Penrose.

Es interesante observar que, una vez más, parece que la biología y la naturaleza ya han descubierto varias formas de desarrollar mecanismos térmicos específicos que promueven la coherencia cuántica y evitan la decoherencia.

Como ya hemos señalado, hay evidencia que muestra que las plantas utilizan rutinariamente el transporte de electrones (con coherencia cuántica) a temperatura ambiente en la fotosíntesis.⁴⁴ ¿Pero que pasa con el cerebro humano? ¿Hay alguna evidencia?

En 2009, Anirban Bandyopadhyay y sus colegas del Instituto Nacional de Ciencia de Materiales en Japón utilizaron nanotecnología para estudiar las propiedades electrónicas y ópticas de los microtúbulos individuales. Descubrieron que los efectos cuánticos pueden ocurrir en los microtúbulos a temperaturas biológicamente relevantes (es decir, lejos del cero), lo que sugiere que la existencia de estados cuánticos

⁴⁴ Publicación original: Hildner, R., Brinks, D. et al. (2013) "Quantum coherent energy transfer over varying pathways in single light harvesting complexes", *Science*, 340 (639): 1448-1451.

coherentes en los microtúbulos a temperaturas cerebrales es una posibilidad real.⁴⁵

Más aún, hay otra pieza curiosa de evidencia en el ámbito de la biología, no de un organismo complejo y altamente desarrollado, sino de un organismo unicelular muy simple llamado 'paramecio'. A pesar de carecer de una sola neurona, este organismo puede moverse, reproducirse y alimentarse. ¿Cómo lo logra? A través del uso de estructuras específicas llamadas... microtúbulos.⁴⁶

Mientras que esta teoría de la conciencia es innegablemente compleja, se esfuerza por abordar dos de los desafíos científicos y filosóficos más profundos del siglo XXI, dilucidar la existencia de la conciencia en los seres humanos y tratar de reconciliar dos teorías aparentemente irreconciliables en una "Teoría del Todo", la mecánica cuántica y la teoría de la relatividad de Einstein.

Con esta propuesta, Penrose ofrece ideas para resolver ambos desafíos simultáneamente. ¿Estás convencido, lector?

⁴⁵ Publicación original: S. Sahu, S. Ghosh, K. Hirata, D. Fujita, A. Bandyopadhyay (2013) "Multi-level memory-switching properties of a single brain microtubule", *Applied Physics Letters*, 102: 123701.

⁴⁶ Original publication: Nakagaki, T., Yamada, H. e Toht, Á. (2000) "Maze - solving by an amoeboid organism", *Nature*, 407: 470.

Después de estas notas introductorias, espero que te resulte más fácil seguir los diálogos con estos cuatro pensadores increíbles con quienes tuve el honor de explorar la naturaleza de la conciencia.



Imagen 12: "Generativo" © Gaia Amoroso

Ya sea que te deje más iluminado o confundido, creo que podemos estar de acuerdo en que una confusión clara siempre es mejor que una certeza poco clara.

DIÁLOGOS I

Conciencia

DIÁLOGOS I

Conciencia

VI. Diálogo con David Chalmers



David Chalmers es profesor de Filosofía y Neurociencia y codirector del Centro para la Mente, el Cerebro y la Conciencia en la Universidad de Nueva York. También es profesor distinguido de Filosofía en la Universidad Nacional de Australia.

Se doctoró en Filosofía y Ciencias Cognitivas en la Universidad de Indiana en el grupo de investigación de Inteligencia Artificial de Douglas Hofstadter.

Chalmers es conocido por formular el "problema difícil" de la conciencia y su trabajo sobre "la mente extendida", la idea de que la tecnología que usamos puede literalmente convertirse en parte de nuestras mentes. Su trabajo sobre lenguaje, metafísica, tecnología e inteligencia artificial también ha atraído mucho interés.

Es cofundador y expresidente de la Asociación para el Estudio Científico de la Conciencia y es codirector de la fundación "PhilPapers".

Es autor de *Reality+: Virtual Worlds and the Problems of Philosophy* (2022), *The Conscious Mind* (1996), *The Character of Consciousness* (2010) e *Constructing the World* (2014).

Más información: <https://consc.net/>

Pregunta: En tu libro *The Conscious Mind* (1996), introduces el "problema difícil de la conciencia" en contraste con los "problemas fáciles de la conciencia," estos ligados a los correlatos neuronales de la conciencia y sus implicaciones para el comportamiento. Después de todos estos años, ¿sigues considerando que el problema de la conciencia es tan "difícil" como lo era entonces?

David Chalmers: Diría que estamos avanzando en la comprensión del problema de varias vertientes, por lo que podemos afirmar definitivamente que hay progreso en el estudio de la naturaleza de la conciencia. Dicho esto, no creo que alguien haya resuelto aún el problema difícil, y creo que sigue siendo un problema fundamentalmente muy difícil de resolver. La distinción básica que hago es entre los problemas "fáciles" de la conciencia, que son aquellos que consisten en explicar varias funciones conductuales y cognitivas, para las cuales tenemos un paradigma explicativo, y el problema "difícil", que es el problema de explicar la experiencia subjetiva, para el cual no tenemos el mismo paradigma, pues parece que los métodos estándar actuales de la ciencia cognitiva dejan abierta la pregunta de por qué toda nuestra biología debería dar lugar a experiencias subjetivas conscientes en primera instancia.

Además, considero que esta distinción básica y relevante sigue ahí, en la medida en que, si se ofrece

una explicación común de la ciencia cognitiva, no resolverá el problema difícil. Así que necesitamos algo nuevo.

Dicho esto, la ciencia de la conciencia se ha desarrollado muy bien sin tener que resolver el problema difícil, por ejemplo, buscando correlatos neuronales de la conciencia, o incluso buscando teorías de la conciencia como la TII (Teoría de la Información Integrada), que no intentan reducir la conciencia, sino conectarla con propiedades físicas y el cerebro.

En última instancia, puede ser que lo mejor que podamos hacer para el problema "difícil", sea algo así como un esbozo de los principios fundamentales que conectan la conciencia con los procesos físicos, y tal vez teorías como la TII estén tratando alcanzar esos principios. No creo que ninguna teoría haya logrado el tipo de evidencia y consenso necesarios para ser realmente aceptada como una teoría de la conciencia.

Mientras tanto, varios científicos y filósofos han estado explorando muchas ideas diferentes que se comprenden mucho mejor hoy que hace 30 años, ya sea el pansiquismo, el enfoque de la mecánica cuántica o la teoría del ilusionismo. Creo que se ha logrado un progreso importante en cada una de estas áreas, pero fundamentalmente, creo que el problema "difícil" es tan difícil como siempre lo fue y continuara siendo.

Pregunta: ¿Cómo concebirías una ciencia rigurosa de manera que se considere un enfoque sólido para la naturaleza de la conciencia? ¿Estaría esto relacionado con tener una fuerte validez matemática y formal? ¿O crees que también sería necesario combinar datos empíricos y evidencias con pruebas matemáticas? Es decir, ¿qué tipo de criterios consideras relevantes para desarrollar una ciencia seria de la conciencia?

David Chalmers: Veo la ciencia de la conciencia como una integración de datos de tercera persona, descripciones objetivas de sistemas cognitivos, con datos de primera persona, es decir, el tipo de datos que podemos obtener desde la experiencia subjetiva.

Podemos obtener datos objetivos a través de los métodos estándar de la ciencia contemporánea, como la observación y la medición, especialmente del comportamiento y del cerebro. Y podemos obtener datos de primera persona al considerar la experiencia subjetiva. Así, lo que queremos de una ciencia de la conciencia es que pueda alcanzar, en última instancia, principios que conecten el reino de la primera persona con el reino de la tercera persona, formulados de manera rigurosa.

Esto exigirá que una ciencia de la conciencia sea realmente rigurosa en varias etapas. En primer lugar, serán necesarios métodos para recolectar datos de primera persona que sean tan rigurosos como nuestros métodos para recolectar datos en tercera persona y,

aunque varios investigadores han reflexionado bastante sobre este punto, nuestros métodos para alcanzar los datos de primera persona son aún muy primitivos.

También necesitaremos métodos para formular la estructura de la conciencia en términos rigurosos. Steven mencionó las matemáticas, aquí es donde considero que las matemáticas pueden tener relevancia. Podemos, por ejemplo, trabajar para encontrar descripciones matemáticas de la estructura de la conciencia. Y puede ser que, al menos parcialmente, principios matemáticos conecten la estructura de los procesos físicos, es decir, datos de tercera persona, con la estructura de la experiencia subjetiva, es decir, datos de primera persona.

Creo que podemos encontrar una aproximación de tal enfoque en la Teoría de la Información Integrada: creo que es una metodología muy prometedora para una ciencia de la conciencia. No obstante, debo ser claro al afirmar que una descripción matemática de la conciencia no agotará toda la naturaleza de la conciencia. El experimento mental de "Mary, la brillante neurocientífica," busca mostrar un poco este punto. Mary podría conocer la estructura matemática de la conciencia de antemano, pero la experiencia de, ver por primera vez el color rojo es una experiencia consciente nueva, que no se podía reducir a todo el conocimiento que ella ya poseía sobre ver colores.

Así, creo que la estructura matemática de la conciencia es, en realidad, al menos una caracterización parcial muy buena de la conciencia. Y esto puede proporcionar lo que algunas personas llaman "fenomenología objetiva", que puede desempeñar un papel crucial en el desarrollo de la ciencia de la conciencia.

Pregunta: En relación con su nuevo libro *Reality+: Virtual Worlds and the Problems of Philosophy* (2022), presenta la "hipótesis de la simulación" que, muy brevemente, defiende que los escenarios presentados por la ciencia ficción como la película "The Matrix" pueden tener una mayor plausibilidad de lo que parecen tener inicialmente, es decir, que tal vez podríamos estar, en este momento, viviendo en un mundo simulado. Esta simulación podría haber sido creada por una civilización ancestral que fue capaz de desarrollar este tipo de tecnología con la capacidad de simular mundos enteros. Centrándose en esta idea de simulación, argumenta que los seres simulados pueden ser tan conscientes como nosotros y me gustaría saber por qué cree que esto es así.

David Chalmers: Podemos volver parcialmente a la antigua pregunta de si necesitamos la biología (basada en el carbono) para producir conciencia, o si es más una cuestión de procesamiento de información, de la computación misma y de la organización funcional del sistema. Siempre he estado del lado de los filósofos y

científicos que afirman que lo que realmente importa para la conciencia no es la biología específica, sino algo más como la organización funcional y/o el procesamiento de la información.

La forma en que planteé este debate en mi libro *The Conscious Mind* en los años 90 fue imaginar esta experiencia de pensamiento de reemplazar gradualmente las neuronas por chips de silicio y argumentar que, si hacemos esto gradualmente, es decir, si reemplazamos neurona por neurona hasta obtener un cerebro completamente constituido por chips, y si esto preserva el funcionamiento normal de la organización de ese sistema, entonces tal cerebro artificial podría preservar la consciencia.

No obstante, no considero que este tipo de métodos nos llevaría a crear simulaciones, sino que nos llevaría al menos a isomorfos de silicio en los que estaríamos conscientes como somos ahora, solo que con biología. Al llegar a este nivel de desarrollo tecnológico, creo que no será necesario mucho más para crear seres simulados que sean igualmente conscientes que nosotros.

¿Porqué? Bueno, lo que defiendo es que una simulación de mi cerebro será básicamente análoga a un isomorfo de silicio de mi cerebro: tendrá muchas partes interactivas que están procesando exactamente la misma información. Así, si lo que realmente importa es la estructura o el procesamiento de la información, en

principio podríamos estar ante una simulación lo suficientemente buena del mundo real.

Sin embargo, es cierto que continuaría existiendo el problema difícil de la conciencia, por lo que no podemos afirmar que realmente comprendemos cómo podría ser consciente una simulación. Sin embargo, no creo que esto sea un argumento concluyente, porque tampoco comprendemos cómo puede ser consciente un cerebro. Lo que argumento, en última instancia, es que el problema de la conciencia es igualmente difícil en el caso de la simulación como en el caso del cerebro humano y que, en principio, si uno puede ser consciente, no veo por qué el otro no pueda serlo también.

Pregunta: Interesante. Sin embargo, diría que incluso el procesamiento de la información realmente depende de la "parte" material del sistema, es decir, que hay algo verdaderamente único y especial en el tipo de biología orgánica que el proceso evolutivo ha desarrollado y que sirve de base para la experiencia consciente en los seres humanos. Pero usted profesor parece pensar que eso no es así, ¿verdad? ¿Que el material que constituye este sustrato no es relevante, siempre y cuando mantenga su capacidad funcional?

David Chalmers: Entiendo. Creo que el sustrato puede marcar una diferencia significativa en la forma en que

se procesa la información, y estoy seguro de que la estructura de las neuronas, por ejemplo, genera una gran diferencia en la forma en que se procesa la información en el cerebro. Dicho esto, no veo por qué eso, en principio, no podría ser simulado. Cualquiera que sean las propiedades idiosincráticas de las neuronas son, me parece, hasta donde puedo afirmar, que no hay nada que sea *incalculable* o *no simulable*.

Ahora bien, si resulta que el Sir Roger Penrose tiene razón, entonces podría haber algunos procesos mecánicos cuánticos especiales que ocurren en las neuronas que no pueden ser simulados en una computadora clásica. Y si ese fuera el caso, entonces necesitaríamos hacer más trabajo.

Pero incluso si Sir Roger está en lo cierto, todavía me pregunto si tal vez podría haber algún tipo nuevo y especial de simulación cuántica, dado que tenemos computadoras cuánticas que básicamente exploran ciertas propiedades físicas de la mecánica cuántica.

Si el premio nobel de Física tiene razón, quizás las cosas vayan mucho más allá: la gravedad cuántica implicaría nuevos tipos de procesos que no pueden ser simulados, incluso en un computador cuántico común. Si ese fuera el caso, podríamos, aún así, construir nuevos computadores de gravedad cuántica que puedan explorar el tipo especial de computación de Penrose y, en principio, no veo por qué no podríamos construir una simulación de un cerebro en uno de estos

nuevos computadores especiales de gravedad cuántica.

Ese tipo de simulación iría más allá de la simulación en una computadora clásica, pero creo que aún sería un enfoque interesante para la simulación.

Pregunta: Sabemos que el problema de la conciencia ha sido parte de la filosofía durante cientos y hasta miles de años, pero la contribución de los filósofos no ha sido muy prolífica. ¿De qué manera pueden los filósofos trabajar para encontrar una solución al problema difícil de la conciencia? ¿Cuál puede ser el papel real de la filosofía para resolver el problema difícil de la conciencia?

David Chalmers: ¿Qué puede hacer la filosofía para resolver el problema difícil de la conciencia? Esa es una buena pregunta: creo que los filósofos han explorado diferentes enfoques para abordar el problema difícil. En los últimos 20 años, hemos visto muchos trabajos interesantes sobre el panpsiquismo con propuestas que intentan abordar lo que hace que el panpsiquismo sea realmente difícil de defender, lo que se conoce como el "problema de la combinación", por filósofos como Philip Goff, Hedda Mørch y Galen Strawson.

Otra teoría interesante desarrollada es la teoría del ilusionismo sobre la conciencia, desarrollada por científicos cognitivos, pero que realmente ha sido

impulsada por muchos filósofos, incluyendo a Dan Dennett, Keith Frankish y otros.

Por lo tanto, los filósofos han intentado realmente buscar soluciones posibles para, al menos, aspectos particulares del problema difícil de la conciencia. Sinceramente, no me importa mucho si es un filósofo o un científico quien logre resolver el problema, pero sospecho fuertemente que tal resolución deberá involucrar algún tipo de interacción entre filosofía y ciencia.

Pregunta: Siguiendo a Sir Roger Penrose, ¿cree que necesitamos un nuevo "tipo" de física para aproximarnos al problema difícil de la conciencia? ¿Qué tipo de teorías podrían tener éxito para abordar la naturaleza de la experiencia subjetiva?

David Chalmers: ¿Qué tipo de teoría podría explicar la conciencia? Me gusta la idea de la matemática de la conciencia, donde podemos mapear la estructura de la conciencia y, en un segundo nivel, mapearla en los procesos físicos del cerebro. Mi opinión es que una teoría de la conciencia puede requerir leyes psicofísicas que conecten los procesos físicos con la conciencia y, finalmente, leyes psicofísicas fundamentales.

Podría ser, por ejemplo, el panpsiquismo, pero no necesariamente tiene que ser el panpsiquismo: un dualista también puede creer en leyes psicofísicas

fundamentales. Giulio Tononi puede ser visto como alguien que propone una teoría de la conciencia, al igual que Sir Roger Penrose. En el fondo, creo que la cuestión de cuál de estas teorías es la correcta dependerá mucho del desarrollo de la ciencia. Alguien podría objetar, en este punto, que una teoría puramente matemática de la conciencia todavía estaría sujeta al problema de "Mary".

Así, argumentaría que debemos distinguir entre el carácter cualitativo de la conciencia y su carácter estructural, lo que nos podría permitir alcanzar leyes psicofísicas matemáticas objetivas que pueden explicar el carácter estructural de la conciencia. Claro, aún tendrá que haber algún trabajo en relación con el carácter cualitativo de la conciencia, como la experiencia de la rojez del rojo.

Pregunata: Algunos neurocientíficos y filósofos sostienen que la conciencia no es más que un error de categoría, un concepto que ya no tiene sentido en nuestro vocabulario científico y filosófico. ¿Por qué deberíamos preocuparnos por el problema de la conciencia en primer lugar? ¿Qué razones pueden ofrecerse para considerar este problema filosófico realmente relevante?

David Chalmers: No puedo convencer a nadie de que se preocupe por algo en específico: hasta cierto punto,

eso depende de cada persona. Hay muchos temas fascinantes en el mundo y no todos tienen que estar interesados en el problema difícil de la conciencia. ¿Por qué es este interesante e importante?

Si estamos interesados en prever el comportamiento de otras personas, entonces, para ese propósito, quizás no sea necesario esclarecer nada fundamental sobre la conciencia. Eso dependerá de muchas cuestiones complicadas, como si el dualismo interaccionista es verdadero, por ejemplo.

Sin embargo, creo que nos importa mucho más que solo el comportamiento de los demás: nos importan las otras personas (¡punto final!) mucho más que simplemente su comportamiento.

Creo que, para mucha gente, es precisamente por el hecho de que las otras personas son conscientes que nos importan. Por eso, para cuestiones éticas y políticas como "¿Cómo debemos tratar a los animales?", considero que es absolutamente crucial saber si los animales son conscientes y qué tipo de estados conscientes poseen: por ejemplo, "¿pueden sentir dolor como nosotros?"

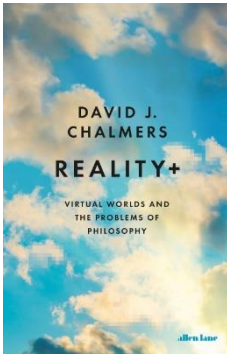
En este punto, alguien podría afirmar: "Puedo saber todo sobre el animal simplemente observando su comportamiento". De acuerdo, pero eso no me dice lo más crucial. Saber cómo debo tratar moralmente a los

animales, lo que implica asumir algo sobre su experiencia subjetiva consciente.

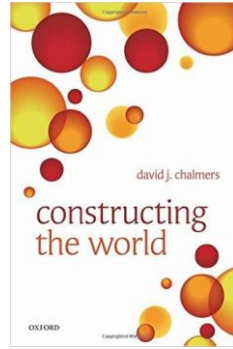
Si consideramos que la experiencia subjetiva está realmente en la base del valor y el significado de nuestras vidas, entonces tenemos al menos una razón para preocuparnos por la naturaleza fundamental de la conciencia.

Podemos también encontrar razones puramente intelectuales. La conciencia es increíblemente interesante, es una anomalía en nuestra imagen del universo, no tenemos un enfoque teórico para entenderla y sin una teoría de la conciencia, nunca podremos tener una teoría completa del universo. Creo que estas son razones más que suficientes para preocuparnos por la conciencia.

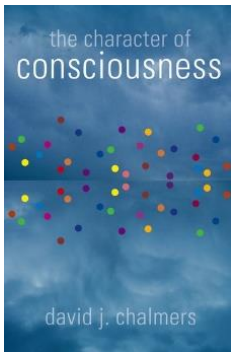
Libros de David Chalmers



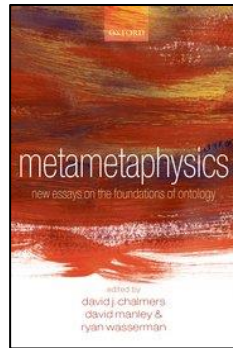
2022



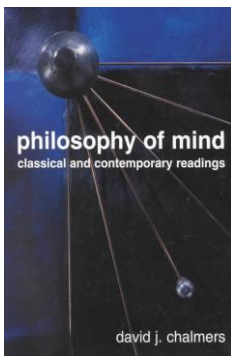
2012



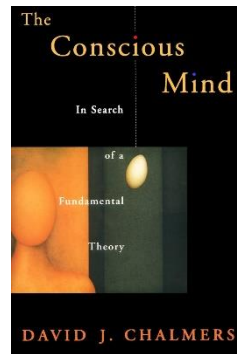
2010



2009



2002



1996

VII. Diálogos con Susan Blackmore



Susan Blackmore es escritora, conferencista, locutora y profesora visitante en la Universidad de Plymouth, Reino Unido.

Sus intereses de investigación incluyen los memes, la teoría evolutiva, la conciencia y la meditación.

La profesora Blackmore escribe para varias revistas y periódicos, bloggea para *Psychology Today* y es colaboradora frecuente y presentadora de radio y televisión.

Es autora de más de quince libros, sesenta artículos académicos y alrededor de ochenta contribuciones en libros.

Algunos de sus libros son: *Dying to Live: Near Death Experiences* (1993), *The Meme Machine* (1999), *Conversations on Consciousness* (2005), *Zen and the Art of Consciousness* (2011), *Seeing Myself: What Out-of-body Experiences Tell Us About Life, Death & the Mind* (2017) e *Consciousness: An Introduction* (2018).

Más información: <https://www.susanblackmore.uk/>

Pregunta: La profesora Blackmore fue una de las primeras científicas en otorgar una mayor relevancia filosófica a los estados alterados de conciencia. Comencemos por su "loca" experiencia que le sucedió cuando era más joven. La profesora tuvo una especie de experiencia "fuera del cuerpo" y eso influyó mucho en su interés por la conciencia en general. ¿Podría contarnos un poco sobre esa historia personal y el impacto que tuvo en sus investigaciones?

Susan Blackmore: Todo ese episodio duró alrededor de dos horas y media. Estaba sentada en el cuarto de un amigo, en mi primer período en Oxford; tenía 19 años, estamos hablando de hace 50 años. Recuerdo que estaba muy cansada, sin dormir durante algunos días, y que había fumado un poco de cannabis, no lo suficiente para explicar todo el fenómeno, aunque posiblemente contribuyó al mismo.

En determinado momento, comencé a descender por un túnel. Era como un túnel de árboles, con hojas alrededor, en dirección a una luz. Todo esto sucedió antes de que se inventara el término "experiencia cercana a la muerte".

En ese período, no tenía ningún conocimiento sobre este tema. El túnel que estaba experimentando se transformó en una experiencia fuera del cuerpo, parecía estar fuera de mi cuerpo, podía mirar hacia abajo y ver mi propio cuerpo allí estirado en el sofá. Durante este tiempo, aún era comunicativa; un amigo

estaba intentando conversar conmigo y me preguntaba. "¿Qué estás viendo ahora?" y en ese preciso momento comencé a viajar, lo que parecía ser viajar por el mundo, y ver todo tipo de cosas.

Para resumir sucintamente la historia, intenté volver a mi cuerpo, pero fue muy difícil: me sentí volviéndome cada vez más pequeña, encogiéndose y encogiéndose hasta quedar muy diminuta. Me sentí muy asustada y de repente empecé a crecer cada vez más, expandiendo la experiencia hacia algo que se transformó en una experiencia mística clásica, de la cual, una vez más, no sabía nada.

Con todas estas fases de experiencias diferentes, sentí que mi "yo" ahora era un "todo". Ya no había un "yo" particular, una Susan, sino algo que parecía ser todo. Muchas (¡otras!) experiencias ocurrieron a lo largo de este episodio, pero eventualmente quedé exhausta y mi amigo me preguntó. "¿Hay algo más?" y pensé: "No, porque soy Todo, ¿cómo podría haber otra cosa?". Los pensamientos seguían ocurriendo, pero no a través de palabras. Y entonces tuve una especie de percepción de que siempre hay algo más.

Necesité dos días para volver a sentir que estaba dentro de mi cuerpo. Esta experiencia me hizo creer en todo tipo de cosas psíquicas y místicas y en cosas de otro mundo. Tal fue, claro, ilógico, pero comprensible en esa época para una joven de 19 años que no sabía nada sobre la naturaleza de estos fenómenos. Este episodio

ocurrió a principios de la década de 1970: no había neurociencia para darnos algunas respuestas. A partir de este episodio, comencé a creer en la telepatía, clarividencia, psicocinesis, fantasmas, poltergeists, ¡todo! Mundos del más allá, vida después de la muerte, almas, espíritus, etc.

A causa de esta experiencia, decidí no aceptar un doctorado prestigioso que me fue ofrecido en una universidad de renombre, y en su lugar, opté por realizar mi propio doctorado sobre lo "paranormal". Me llevó cerca de cinco años de investigación sobre todos estos temas paranormales para descubrir que es casi seguro que no existen en absoluto y que son una creación de nuestra mente.

Esta es la breve historia de cómo me transformé de una creyente de todo tipo de cosas extrañas y después decidir que las verdaderas preguntas no son: ¿Existe un espíritu, o un alma, o una conciencia más allá del cerebro?. Sino que la verdadera cuestión está aquí y ahora ¿Qué experiencia fenomenal es esta que tengo? ¿Cómo es que un cerebro crea esta experiencia? ¿Cómo contribuyen un cerebro, un mundo y un cuerpo a esto? ¿Cuál es la naturaleza de la conciencia? Estas son las verdaderas preguntas, que son mucho más emocionantes que si existe vida después de la muerte, espíritus o cualquier otra cosa.

Pregunta: Siguiendo sus experiencia, sabemos que muchos tipos de drogas pueden inducir los llamados estados alterados de conciencia. ¿Cree que podemos usar, por ejemplo, psicodélicos como el DMT o el LSD para estudiar la naturaleza de la mente consciente? ¿Cree que esas metodologías son útiles desde el punto de vista de comprender el fenómeno de la subjetividad, dadas las varias dificultades metodológicas que existen actualmente en la investigación científica de la conciencia?

Susan Blackmore: Eso depende de a que te refieras con la "naturaleza" de la mente. Si creemos que estos métodos nos van a proporcionar respuestas rápidas, entonces no considero que sean realmente útiles. Pero desde una perspectiva más general, sí, pueden tener alguna utilidad.

También puede depender de la relación personal que se tenga con esas sustancias, y si tal relación puede permitir ganar alguna comprensión a través de sus propias vivencias o si la investigación que se está realizando nos ofrece alguna explicación. Sin embargo, diría que la mente cambia de forma tan radical durante un episodio de influencia de sustancias que no podemos aprehender algo sustancial de inmediato.

También puede darnos alguna luz acerca de que un estado normal de conciencia no es el único, por ejemplo, piensa en lo que decía William James en 1890, sabemos que, más allá del velo de esta experiencia

particular que estamos teniendo, existen innumerables otras maneras de estar consciente que están muy cerca de esta, pero necesitamos algo como una droga para llevarnos hasta allí, o una experiencia mística espontánea.

Creo que una de las investigaciones más interesantes que he leído sobre este tema es un artículo reciente de Timmermann y colegas (2019)⁴⁷ que es el ingrediente psicoactivo de la ayahuasca, que perturba las principales redes funcionales del cerebro.

En el cerebro encontramos varias redes principales, pero una de ellas en particular, la red de modo por defecto (Default Mode Network, DMN), es de hecho la que subraya el sentido de uno mismo. La DMN es la red de largo alcance que integra el esquema corporal con los recuerdos sobre uno mismo, sus opiniones, etc.

La DMN se conecta a la unión temporoparietal derecha y es en esta área del cerebro donde el esquema corporal se conecta con los sistemas de control en la corteza frontal y con la memoria en los lóbulos temporales, entre otros. Esta red se interrumpe cuando se consume DMT (dimetiltriptamina). Al mismo tiempo, en otras partes del cerebro, como por ejemplo en la corteza visual, hay un aumento de la actividad local que

⁴⁷ Publicación original: Timmermann, C., Roseman, L., Schartner, M. et al. (2019) "Neural correlates of the DMT experience assessed with multivariate EEG", *Sci Rep*, 9: 16324.

puede explicar las alucinaciones visuales que ocurren con el consumo de DMT.

Cuando el sistema que produce el sentido del "yo" es perturbado, dejamos de sentir que existe un "yo". Se podría decir que esto significa ver a través del poderoso "yo" consciente, el "yo" que consideramos tiene conciencia o libre albedrío.

Para mí, todas estas son ilusiones en mi mente, y el hecho de que el DMT perturbe el "yo" significa que un individuo puede experimentar esto -durante algunas horas con ayahuasca o 15 minutos al fumar DMT-, lo que significa tener esta experiencia sin el sentido "normalizado" del "yo".

Todo este tema me parece relevante para comenzar a ver aspectos relacionados con las ilusiones de la conciencia. Creo que la mayoría de las teorías de la conciencia siguen atrapadas en estas ilusiones y no progresan por confundirlas con algo real.

Tal vez los psicodélicos en general puedan ayudar en este sentido, y creo que toda la investigación que estamos llevando a cabo actualmente está empezando a revelar mucho sobre cómo funciona la mente. Y habrá más por descubrir: ahora que todo ha comenzado y la ley ya no puede frenar esta línea de investigación (en el pasado hubo leyes restrictivas contra el uso de drogas en investigación científica), creo que aprenderemos mucho.

Pregunta: En el ámbito de la exploración de la conciencia, ¿Y qué hay de los sueños lúcidos? Sabemos que están mucho más cerca de estar despiertos que en un estado de sueño profundo, ¿Cómo percibes las posibles contribuciones de las investigaciones empíricas sobre estos sueños? ¿Pueden tales estudios proporcionar información valiosa y ayudar en el mapeo y comprensión exhaustiva de los diversos estados de conciencia?

Susan Blackmore: Cuando Steven afirma que el sueño lúcido es un estado similar a estar despierto, no estoy seguro de que sea realmente el caso. Estos estados ocurren principalmente en la fase REM del sueño, y eso significa que necesitamos estar durmiendo adecuadamente. Es en esta fase REM cuando podemos tener sueños lúcidos.

Estos estados ocurren en los periodos más activos del REM, donde hay más actividad que en los periodos de menor actividad del REM. Pero lo que es realmente interesante es que – y, una vez más, se relaciona con la red de modo por defecto (DMN) – las conexiones entre la unión temporoparietal y los lóbulos frontales se vuelven más fuertes en los sueños lúcidos, lo que sugiere que esta es una base psicológica para la sensación de “ah, estoy aquí, ahora, y puedo controlar el sueño”.

Una vez más, me encanta descubrir estas cosas, porque simplemente destruyen el tipo de teorías alternativas

sobre lo que sucede en un sueño lúcido, como el despertar de nuestras almas y todo tipo de cosas absurdas que la gente imagina. Lamentablemente, la psicología convencional aún no está realmente enfocada en la relevancia científica de los sueños lúcidos y actualmente no encontramos suficientes investigadores trabajando en este tema. Considero que estos estados son, sin lugar a dudas, una fuente de aprendizaje muy relevante sobre nuestros cerebros y nuestras mentes.

Pregunta: Con su extensa experiencia de alrededor de 30 años en la meditación Zen, ¿Cómo ha influido esta práctica en su perspectiva sobre la mente consciente? ¿De qué maneras cree que la meditación, con su enfoque en la atención plena y la autoconciencia, proporciona conocimientos sobre la naturaleza de la conciencia? Además, considerando los aspectos subjetivos e introspectivos de la meditación, ¿Cómo enfrenta los desafíos de incorporar tales conocimientos en nuestra comprensión de un estudio científico más amplio de la conciencia?

Susan Blackmore: Depende de lo que entiendas por comprensión, en el sentido de que no estoy segura de que la meditación tenga ese propósito. Creo que su objetivo principal es aclarar la mente y esa es la mejor razón para querer meditar. En cierto modo, para mí, me ha ayudado a entender cosas sobre la mente, también

me ha ayudado a soltar las ilusiones, dejar de creer que hay un yo aquí dentro y un mundo allá afuera, dejar de ansiar ser algo más importante.

Lo que sabemos es que la mente de quien medita cambia de forma bastante radical en el sentido de ser más abiertos a todo en el mundo, más receptivos, menos apegados, y para eso es que la meditación sirve para mí, aunque ese cambio sea muy lento. ¿Puede esto ayudarnos a comprender la mente? Bueno, sí, claro, porque los practicantes de meditación a largo plazo tienen una enorme eficacia en entrar en diferentes estados de conciencia, y podemos crear investigaciones interesantes a partir de eso.

En los últimos años he practicado la Meditación Jhanas, en la cual se supone que hay ocho estados discretos de conciencia que pueden ser alcanzados mediante la concentración profunda. Escuché hablar sobre esto hace décadas y pensé en lo increíble que sería que la existencia de estados alterados específicos pudiera ser alcanzada simplemente siguiendo ciertas instrucciones. Pensé que nunca sería capaz de alcanzar esos estados mediante la meditación. Pero gracias a un gran maestro de Jhanas que conocí en Inglaterra, he tenido el privilegio de participar en sus retiros.

He descubierto que parece ser realmente posible simplemente siguiendo algunas instrucciones (bastante intensas) alcanzar algunos de estos estados: comenzando con una concentración profunda y luego

realizando varias otras acciones, podemos entrar en estados alterados claramente definidos que podemos controlar nosotros mismos. Hasta ahora, solo he logrado ingresar a los tres primeros estados, y quizás un poco del cuarto, y continúo practicando con el objetivo de alcanzar otros estados.

Por la práctica que he realizado, estoy convencida de que esos antiguos pueblos que practicaban estas meditaciones realmente descubrieron estos diferentes estados de conciencia y también descubrieron cómo alcanzarlos. Esto nos informa sobre lo que la mente es capaz de hacer: cuando puede lograr esto, o incluso simplemente al leer y comprender lo que las personas están relatando sobre esas experiencias, entonces sabemos que la mente puede estar en un tipo de relación completamente diferente consigo misma, con el mundo y con la idea que tiene de sí misma. Por lo tanto, sí, creo que la meditación puede aclararnos algunas ideas sobre la naturaleza de la mente.

Pregunta: Tengo la impresión de que estos diferentes estados alterados de conciencia pueden ser reconocidos a través de imágenes neuronales, ¿verdad? ¿Con suficiente práctica, podríamos ver ocho tipos de actividad cerebral correspondientes a las ocho fases diferentes de esta meditación específica?

Susan Blackmore: Según mi conocimiento, solo se han realizado dos experimentos neurocientíficos sobre este tema, ambos con el practicante de meditación Lee Brasington. Estos estudios cerebrales con fMRI o alguna otra técnica de imagen similar en neurociencia son muy costosos.

El problema es que el sujeto debe ser capaz de indicar a los científicos, en este estado de concentración muy profundo, cuándo está cambiando al siguiente estado meditativo y asociar ese cambio a su actividad cerebral. Ciertamente, los neurocientíficos deben haber identificado algunos cambios en el momento en que el sujeto indica que está pasando del segundo al tercer estado, por ejemplo. Pero no me parece que tengamos datos suficientes y no creo que Leigh haya logrado pasar por los ocho estados, porque los dos últimos estados están asociados a lo que se podría llamar "inconsciente".

Son estados que están tan "alejados" que sería muy difícil comunicarlos en ese entorno experimental, pero sería increíble, sin duda. Es muy, muy difícil meditar adecuadamente dentro de un escáner fMRI, con todo ese ruido y todo lo demás. Así que creo que esa previsión teórica es quizás demasiado optimista.

Pregunta: Sabemos que la conciencia es realmente difícil de investigar de forma rigurosa. Sin embargo,

encontramos individuos como Deepak Chopra, con quien la profesora Blackmore ya ha tenido varios debates, que apoyan algunas ideas espirituales bastante dudosas sobre la mente. ¿Qué opina de esas ideas en general?

The investigation of consciousness poses significant challenges, and individuals like Deepak Chopra often promote (dubious!) spiritual ideas about the conscious mind. In your engagements with him and discussions on such topics, what is your perspective on these spiritual ideas? How do you approach the balance between exploring the mysteries of consciousness and maintaining scientific rigor, especially when addressing concepts that may lack empirical support or scientific grounding?

Susan Blackmore: Si alguien está interesado en Deepak Chopra, eche un vistazo a mi Facebook y vea un video que publiqué sobre mi debate con él y verá a un autoproclamado gurú espiritual comportándose de una manera bastante dudosa.⁴⁸

No puedo responder a la pregunta sobre todos esos gurús, pero puedo responder sobre Deepak, quien sostiene que la conciencia es primaria y que la materia no existe. Eso es simplemente absurdo: no tiene ningún significado real y la mayor parte de las cosas que afirma no significan nada.

⁴⁸ Video original: youtube.com/watch?v=_ZFGkqhNhgM.

Los consejos prácticos que da sobre habilidades de cómo entrenar la mente y algunas de las cosas que dice sobre la meditación y sus consecuencias, son precisos, en ese aspecto, realmente tiene un conocimiento interesante. Pero cuando se trata de la relevancia de sus teorías sobre la mente, estas son muy insatisfactorias desde un punto de vista filosófico.

Consideremos el problema filosófico mente-cuerpo: si somos materialistas, defendemos que todo lo que existe en el mundo está constituido por materia, y así no podemos explicar la experiencia subjetiva, siguiendo lo que llamamos el "problema difícil de la conciencia", que no está resuelto. Tal vez, en realidad, sea la pregunta equivocada, el problema equivocado.

Pero si afirmamos que somos idealistas, tenemos que defender que todo lo que existe es conciencia. Pero entonces dejamos de poder explicar la materia. Claramente, ninguna de esas dos posiciones funciona. El dualismo no funciona porque estamos ante dos tipos de cosas completamente diferentes. Así, hay algo profundamente equivocado, y por eso defiendo un supuesto ilusionismo en relación con la conciencia, pero Deepak es solo un idealista, aunque no lo admita.

Él solo dice que todo es conciencia y que la materia no existe. ¡Genial! Esta tesis no hace ninguna predicción, no hay una base teórica para defenderla, no hay conclusiones que se puedan sacar de esta afirmación y, sin embargo, ¡a la gente le encanta escucharlo hablar! A

la gente le encanta porque piensa: “de alguna manera, mi conciencia es tan maravillosa e importante, y puedo hacer todo esto con mi conciencia. Él está realmente consciente, así que debe ser muy espiritual”.

Claro, estoy haciendo una caricatura aquí, pero realmente creo que sus ideas, en términos de lo que él piensa, son, desde el punto de vista científico, simplemente vacías. Y lo interesante es que recurre a alguna investigación científica, pero la distorsiona y hace interpretaciones absurdas para adecuarlas a esa tesis idealista que defiende.

Pregunta: La profesora Blackmore afirmó que es una ilusionista, pero dijo que rechazaría tanto el materialismo como el idealismo. ¿Podemos decir que el ilusionismo tal vez sea una especie de combinación con el materialismo, en el sentido de que el dolor es un material que crea una ilusión?

Susan Blackmore: It is a very good question. Somebody asked me that question before at the psychedelics conference, and all I can say is: most illusionists are materialists, that is for sure. On the other hand, I do not have anything to replace it with.

Es una pregunta interesante. Todo lo que puedo decir es que la mayoría de los ilusionistas son materialistas. Eso es seguro. No me considero materialista en el sentido tradicional, pero, por otro lado, no tengo nada

para sustituirlo. Mi idea de ilusionismo es bastante amplia, todas estas formas existentes de pensar sobre la conciencia están equivocadas y aún no hemos encontrado una salida a este dilema, y eso es lo que llamamos el "problema difícil".

Por otro lado, quienes intentan entender cómo surge la conciencia del cerebro parecen caer en una especie de estructura dualista. Aunque todos estos materialistas, con excepción de David Chalmers, que inventó el término "el problema difícil", siguen siendo materialistas sin realmente resolver el problema. Ellos están afirmando, como yo, que "si pensamos que la conciencia es algo que tenemos, si pensamos que tiene poder y hace algo, si pensamos que la conciencia evolucionó para un propósito, entonces estamos ilusionados".

Pienso que Daniel Dennett y Keith Frankish, por ejemplo, dirían "sí, existe un cerebro material a partir del cual se construyen las ilusiones" y están intentando resolver el problema de "¿Cómo surgen esas ilusiones en la mente?". Así, ellos sustituyen el "problema difícil" por el problema de la ilusión, o el metaproblema, como lo llama Chalmers. Este cambio en el problema central es un progreso relevante desde mi punto de vista.

No soy filósofa, pero si me preguntan, diría que soy una monista neutral. Soy monista porque creo que el dualismo no funciona. Pero no creo que nuestros conceptos actuales de "materia" o "mente" sean útiles.

No sé realmente en qué consiste el universo, quizás alguna física fundamental pueda ayudar, aunque enfrentamos grandes dilemas por no poder combinar la mecánica cuántica con las teorías einsteinianas. También podemos considerar algunas de las teorías más recientes sobre la información, que sostienen que todo lo que existe es información. Y también podemos pensar en la termodinámica y la entropía.

En fin, creo que es en estos enfoques posibles donde alguien podría llegar a una solución: personalmente, estoy tratando de entender todos estos nuevos conceptos, lo cual es un poco difícil debido a la multiplicidad de enfoques y conocimientos involucrados. Pero creo que por eso el tema de la conciencia es el desafío más estimulante de la ciencia actual.

Pregunta: La profesora Blackmore abordó su perspectiva ilusionista sobre los estados conscientes. Sin embargo, yo diría que también puede haber una relación entre la ilusión de la conciencia y la ilusión del "yo". ¿Cuál es la relación entre ellas, si es que existe alguna?

Susan Blackmore: Lo mejor que puedo decir es que son ilusiones extremadamente cercanas. El comienzo de la ilusión de la conciencia es la separación del "yo" del "otro". Es el problema de otras mentes, por así

decirlo; la separación entre mí y otras personas. Es la sensación de que esta es mi conciencia y que la mía es diferente a la de otras personas, porque no puedo saber cómo es para los demás ver el azul y si es igual a mi azul, y todas ese tipo de cosas".

¿Es todo una ilusión? No: diría que las cosas que me interesan en términos de posibles ilusiones son principalmente las trampas.

Si miramos el trabajo de Daniel Dennett, especialmente en su libro "Consciousness Explained", él habla sobre todas las trampas en las que las personas caen, y también realiza trabajos posteriores sobre las "bombas de intuición", y así sucesivamente.

La teoría de los "múltiples borradores" o "versiones múltiples" (multiple draft theory) avanzada por Dennett nos lleva a concluir algo que para la mayoría de los neurocientíficos es completamente extraño: que no hay distinción entre los procesos neuronales que son conscientes y los que no lo son. De hecho, no hay diferencia alguna. No hay ningún hecho al respecto; es insensato afirmar que esos procesos cerebrales que ocurren aquí o allá son los de la conciencia.

Toda la búsqueda de los correlatos neuronales de la conciencia en mi opinión está completamente equivocada. Proviene de estas ilusiones, porque sostiene que la conciencia misma emerge de algún

proceso específico, de alguna área particular del cerebro. Pero según Dennett, simplemente no es así.

Si le preguntara al lector “¿Está consciente ahora mismo en este preciso momento?” -al leer esta pregunta, podría haber tenido una experiencia extraordinaria y pensar: “¡Por supuesto que lo estoy! Pero espera, hace un momento solo estaba leyendo esta frase, pero ahora algo ha cambiado. “¡Me he vuelto más consciente!”

Esto me lleva a pensar que, volviendo a relacionarlo con la teoría de los “múltiples borradores”, si el lector no sabe si estaba consciente antes de pensar en eso, entonces nadie lo sabe. No podemos colocar ningún medidor de conciencia en el cerebro y declarar dónde está la conciencia, cuándo una persona está consciente o no en un momento determinado.

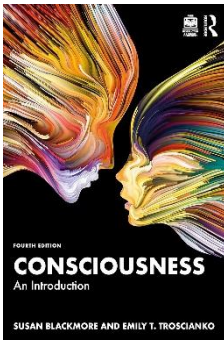
Entonces, según Dennett, si no lo sabemos, no hay ningún hecho relevante al respecto. Todo lo que hacemos es, retrospectivamente, después del hecho, decir que éramos conscientes de esto y no de aquello, siempre es una atribución posterior al hecho mismo.

Volviendo a tu pregunta, ¿Será que el “yo” y la conciencia están basados en la misma ilusión? De cierta manera, sí lo están, porque soy yo quien está consciente, lo cual implica una separación entre mí y esa cosa llamada “mi conciencia”, así como las cosas de las cuales soy consciente. Entonces, supongo que mi respuesta sería “sí” y “no”. Sí, están íntimamente

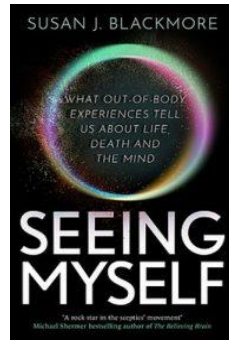
relacionados, pero no podemos separarlos en varios niveles diferentes de ilusión.

Para concluir, pienso que el campo de los estudios de la conciencia está lleno de ilusiones y no haremos ningún progreso real hasta que comprendamos cómo surgen estas ilusiones y aprendamos a "ver" a través de ellas.

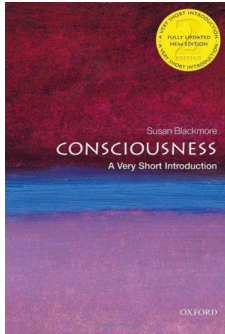
Libros de Susan Blackmore



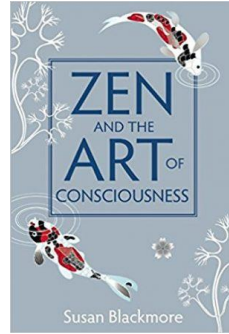
2024



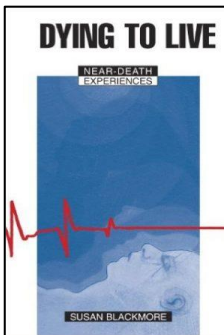
2020



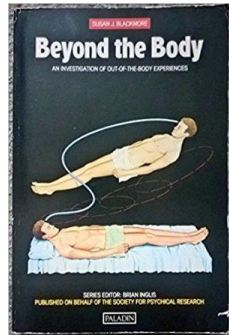
2017



2011



1993



1992

VIII. Diálogos con Nicholas Humphrey



Es un psicólogo teórico conocido internacionalmente por su trabajo sobre la evolución de la inteligencia y la conciencia humana.

Sus intereses son variados, estudió gorilas de montaña con Dian Fossey en Ruanda; fue el primero en demostrar la existencia de la "visión ciega" tras daño cerebral en monos y propuso la célebre teoría de la "función social del intelecto".

Es un autor prolífico y ha publicado varios libros: *Consciousness Regained* (1983), *The Inner Eye* (2003), *Seeing Red: A Study in Consciousness* (2009), *A History of the Mind* (2012), *The Mind Made Flesh* (2012), e *Soul Dust: the Magic of Consciousness* (2012).

Recientemente publicó su nuevo libro: *Sentience: The Invention of Consciousness* (2023).

Más información: <https://www.humphrey.org.uk/>

Pregunta: En el verano de 1966, el Profesor Humphrey conoció a Helen, una macaca ciega cuyo córtex visual había sido removido con fines de investigación científica. ¿Qué aprendió el profesor sobre la conciencia a partir de su experiencia con Helen?⁴⁹

Nicholas Humphrey: Fue una experiencia muy extraña cuando encontré a esta macaca por primera vez. Era estudiante de doctorado en la Universidad de Cambridge y en el laboratorio había una macaca que no había sido operada por mí, sino por otro colega, con el objetivo de descubrir cuál era la función del córtex visual. Entonces, una colega mía removió todo el córtex visual del cerebro de esta macaca y, sin sorpresa, al parecer había quedado completamente ciega.

Conocí a Helen un año después de la operación y ella estaba sentada, mirando vagamente a la nada, no estaba ni un poco interesada en usar sus ojos. Pero me intrigó esto, porque a Helen le habían removido el córtex en la parte trasera del cerebro, pero se había dejado intacto el sistema visual más antiguo, que es el sistema visual usado por peces, ranas y otros vertebrados no mamíferos, el sistema visual del colículo superior del mesencéfalo, que les permite ver perfectamente bien. Una rana puede usar sus ojos y atrapar moscas y así sucesivamente, sin ningún problema.

⁴⁹ Helen en acción: [youtube.com/watch?v=rDIsxwQHwt8](https://www.youtube.com/watch?v=rDIsxwQHwt8).

Entonces, como este sistema todavía estaba intacto en este mono, me puse a pensar si de hecho podría ser entrenado para ser utilizado nuevamente, incluso si un mono con un cerebro "normal" no hace uso de este sistema visual antiguo. ¿Podría Helen recuperar su visión utilizando este sistema del mesencéfalo?

Cuando me enfrentaba a estas preguntas, mi supervisor se ausentó durante una semana para una conferencia en Zúrich, Suiza y aproveché la oportunidad para sentarme con esta macaca y jugar con ella, tratando de hacer que interactuara conmigo de alguna manera. Para mi sorpresa, descubrí que en pocas horas ella estaba usando sus ojos.

Ella claramente estaba prestando atención a lo que yo estaba haciendo. Por ejemplo, puse un trozo de fruta delante de ella para que extendiera la mano y la tomara y ella lo hizo exactamente así. Al final de una semana, ya podía extender la mano para tocar una pequeña luz colocada frente a sus ojos o tocar un objeto que se colocaba en la punta de un palo y se movía delante de sus ojos.

Con estos resultados sorprendentes (dado que no se suponía que Helen observara nada de lo que le estaba colocando frente a los ojos), me entusiasmé mucho y envié un telegrama a mi supervisor en Zúrich, pero no le gustó lo que le había escrito. Le envié un telegrama diciendo "Enseñé a Helen a ver. No lo creerás". En ese momento, yo tenía 20 años y mi supervisor era un

distinguido profesor mayor y no quedó muy satisfecho con eso. Regresó a Cambridge y uno o dos días después, logré convencerlo de que fuera a ver a Helen, y tuvo que admitir que algo sorprendente había ocurrido esa semana: Helen claramente había recuperado la capacidad de usar sus ojos.

Entonces, como resultado de todo este episodio, mi supervisor me permitió seguir trabajando con Helen y terminé trabajando con ella durante unos 7 años. Al final de esos 7 años, ella aparentemente había recuperado su sistema visual normalizado. Corría por la habitación, recogía objetos del suelo sin chocar con obstáculos, e incluso podía extender la mano y capturar moscas que pasaban cerca de ella.

Muchas personas concluyeron que Helen tenía una visión normal, pero esto me intrigó. Cuando la macaca Helen estaba conmigo, relajada y confiada, podía ver, pero si se sentía amenazada o ansiosa, su visión desaparecía y comenzaba a tropezar, como si estuviera de nuevo en completa oscuridad. Parecía que solo podía ver lo que no necesitaba pensar mucho. Si tenía que pensar, su confianza y su capacidad de ver desaparecían.

Pensé que era una forma de visión extraordinaria: era una visión en la cual el propio mono no creía realmente que estaba viendo. Entonces, escribí un artículo donde llamé a esta capacidad "ver y la nada", en eco al libro de Jean-Paul Sartre [Ser y la Nada], ya que sentí que

claramente faltaba algo. Pronto descubrimos qué era lo que faltaba. Inspirado por este episodio, mi supervisor probó a pacientes humanos de una manera diferente y estableció el famoso fenómeno de la "visión ciega en humanos".

Descubrió que un ser humano con graves daños en el córtex visual que creía estar ciego podría, de hecho, usar los ojos para ver en esa área del campo ciego, el paciente podría adivinar lo que estaba sucediendo frente a su campo de visión, deduciendo la forma y la posición de un objeto particular, pero al mismo tiempo afirmaba que "no hay nada ahí", "no entiendo", "esto no tiene sentido", "mi visión no tiene nada que ver conmigo".

Entonces, a partir de eso, planteé la gran pregunta: si un individuo puede ver y una rana puede ver y un pez puede ver sin utilizar el córtex visual y sin tener sensaciones visuales, ¿Entonces para qué sirve el sistema secundario y cuál es su propósito? ¿Cuál es la utilidad de las sensaciones visuales? Y esa es realmente la pregunta en la que he estado trabajando durante el resto de mi vida. Durante los últimos 50 años, he intentado descubrir el papel funcional de las sensaciones físicas.

The upshot of this first experience with Helen was that my supervisor allowed me to continue to work with her and ended up working this monkey for seven years. And, by the end of seven years, she apparently was able

to see everything normally. She would run around her room, picking up objects of the floor, not bumping into obstacles, she could even reach out and capture flies as they passed by her.

But you might think, and many people did think, that it looked like she had a normal vision, but I was puzzled by it: something was not normal about it, because when she was with me, she was relaxed and confident, and then she could see, but if she was threatened or anxious at all, her vision would disappear, she would blunder around, as if she was in the dark again. So, it seemed that she could only see if she did not have to think too hard about it. If she did, her confidence deserted her.

I thought it was a very extraordinary form of vision: it was a vision which the monkey herself does not believe in. So, I wrote a paper which I called this capacity "seeing and nothingness", echoing the Jean-Paul Sartre's book , since I thought that there was something clearly missing. And we soon found out what was missing, because my supervisor went on this time to test human patients in a different way, and he established the famous phenomenon of "blindsight" in humans.

He discovered that a human with major damage to the visual cortex who believed he was blind could, in fact, use their eyes to see in that area of the blind field. They could guess what was up: they could guess the shape and position of an object in the visual field, but all the

time they would say “there is nothing there”, “I do not understand”, “this does not make sense”, “my vision has nothing to do with me”.

Y, claro, esto me llevó a plantear otra gran pregunta ¿Por qué las sensaciones visuales y todas las demás sensaciones tienen las cualidades fenomenales tan extrañas que tienen? ¿Cómo es ver el rojo, o sentir el sabor del azúcar, o escuchar un grito o el sonido de las campanas? No solo estamos obteniendo información sobre el objeto. Estamos obteniendo información en una dimensión diferente, en la dimensión fenomenal, sobre cómo se relaciona con nosotros, sobre cómo es para nosotros recibir esa información.

Pregunta: El profesor Humphrey afirma que la conciencia es un producto de la evolución porque confiere beneficios específicos de supervivencia a la humanidad...

Nicholas Humphrey: No digo que sea un producto de la evolución porque confiere ventajas de supervivencia. Lo que pregunto es ¿Qué más puede ser, sino un producto de la evolución? Al menos en este caso, debe conferir ventajas de supervivencia. Entonces, tenemos que descubrir cuáles son esas ventajas exactamente...

Pregunta: Pero también sostiene que la conciencia es una forma de ilusión que nuestro cerebro crea para nosotros mismos. Desde ese punto de vista, la conciencia se percibe como una construcción mental, una representación interna de la realidad que experimentamos.

Nicholas Humphrey: Solía afirmar eso, sí. Pero eso sugiere que la conciencia es una ilusión. He leído varios artículos e incluso otros libros que describen la conciencia como una ilusión y ahora se ha convertido en una visión popular entre ciertos filósofos, como Keith Frankish y Daniel Dennett, quienes son considerados ilusionistas de la conciencia. Sin embargo, rápidamente me alejé de esa etiqueta.

No creo que sea la manera correcta de describir la conciencia. Pienso que llamar a algo una ilusión sugiere que es un error, que estamos equivocados al atribuir las cualidades que hacemos a la experiencia. No creo que sea así: creo que cuando vemos el color rojo o olemos una rosa, esto es realmente como parece ser. Es realmente así como sentimos esa experiencia subjetiva. Así es como es, y describirlo como una ilusión es subestimar el papel que desempeña en nuestra psicología y nuestra fenomenología.

Para mí, la conciencia fenomenal es una descripción verídica de cómo es para un ser humano tener esas experiencias. Y la cuestión central es: “¿por qué representamos esto de esa forma?”. Las sensaciones

son representaciones, podrían haber sido simplemente representaciones de los hechos, simples hechos físicos, pero no lo son. Son representaciones de cómo nos sentimos al recibir esos estímulos a través de nuestros cuerpos.

Pregunta: Es interesante seguir ese pensamiento porque, en la actualidad, se considera un "surrealista" en relación a los "qualia" o a la "conciencia". ¿Puede explicarnos qué pretende defender con esa posición?

Nicholas Humphrey: Estaba tratando de encontrar otro término para describir mi posición. Hubo un artículo escrito y publicado sobre ilusionismo por el filósofo Keith Frankish. En mi respuesta a ese artículo, defendí que "No, creo que no es una ilusión; no es irreal; de hecho, es súper real". La rojez fenomenal es más roja que el rojo; el dolor fenomenal es más doloroso que el dolor. Y tomé esa frase de Pablo Picasso, quien fue uno de los primeros surrealistas, aunque no aceptara el término. Picasso había producido una escultura famosa de una cabra y afirmó: "mi cabra es más cabra que cualquier cabra real".

En otras palabras, Picasso intentaba expresar en su arte la esencia del objeto que creaba. Él decía que iba más allá, que era más profundo que la propia realidad física. Y creo que tal vez esa sea la manera más interesante de hablar sobre las sensaciones fenomenales y la

conciencia. En cierto sentido, parecen ir más allá de los hechos superficiales de la realidad.

Pregunta: El profesor Humphrey también sostiene que existe una estrecha relación entre los "qualia" y lo que podemos llamar el "yo fenomenal". ¿Por qué cree que existe una relación cercana entre estos dos conceptos?

Nicholas Humphrey: Pienso que los qualia, la experiencia fenomenal, son la base del "yo". Y, de hecho, ese es el camino a seguir en nuestra economía psicológica y es por eso que ha evolucionado. Lo que hacen las sensaciones es dar una realidad sustancial a nuestro sentido de quiénes somos y de nuestra existencia en el mundo.

Notoriamente, David Hume, el filósofo, afirmó que, cuando intenta examinar su propia mente y descubrir lo que significa ser él mismo, no encuentra nada más que sensaciones. Y se sintió decepcionado con esto, ya que las sensaciones son efímeras, no parecen tener ninguna realidad continua y no existe una base sustancial para el "yo", basada solo en las sensaciones.

Creo que tal visión está completamente equivocada. Las sensaciones de hecho nos dan la base más sólida que podríamos tener para nuestra existencia en el mundo. Son evidencias siempre presentes de cómo vivimos nuestras vidas y de cómo somos importantes, de cómo es ser nosotros mismos y, lo más importante,

de nuestra propia individualidad. Las experiencias que tenemos son diferentes de todo en el mundo y de las de otras personas.

Es obvio que podemos asumir que otros humanos tienen experiencias como las nuestras. Pero la evidencia no está ahí. Todo lo que sabemos es que mi sensación de rojo se asemeja a la sensación de rojo que tuve antes. Parece tener algo en común con mis otras sensaciones y otras modalidades. Todo está hecho a mi estilo, pero esa es la prueba de mi realidad continua.

Describo la continuidad de las sensaciones como semejante a la continuidad de las obras de arte de un determinado pintor. Y así como podemos identificar el estilo de Cézanne, Vermeer o Picasso en sus obras, todas mis sensaciones son mías, y eso me da una realidad en el ámbito de las cosas que encuentro muy significativa. Cuando crecemos, descubrimos que eso es lo que constituye nuestro centro psicológico, y entonces asumimos que otros humanos tienen un centro del "yo" equivalente, en el cual el juego de sensaciones es similar al nuestro.

Para cada uno de ellos, ese juego será individual, privado e importante. Una vez que adoptamos esa visión de las otras personas, los parámetros de la vida social comienzan a cambiar. Pasamos a vivir en lo que he llamado "La sociedad de los yo", de los "yo" fenomenales. Y esa es ciertamente la base de la cultura humana y de la civilización humana. La gran pregunta

es si eso va más allá de los humanos y si otros animales también pueden ser considerados como poseedores de un "yo" fenoménico.

Pregunta: Siguiendo esa idea, el profesor Humphrey también sostiene que la conciencia tiene un papel específico relacionado con proporcionar una especie de significado a la vida y a la existencia humana en general. ¿Podría elaborar un poco más sobre este aspecto específico de la conciencia, que a menudo es ignorado por los pensadores que se dedican a estudiar este fenómeno?

Nicholas Humphrey: Esta es una gran cuestión, dar sentido a la vida, sí. Bueno, ¿Por dónde empezamos a encontrar significado? Primero, nos da una noción de nuestra propia importancia, de que no somos fenómenos puramente físicos, de que existimos en otro plano o tenemos una dimensión espiritual en nuestras vidas. De cierta manera, existimos fuera del tiempo y del espacio, en un reino que no puede ser descrito por las cualidades materiales de la materia física.

Este es un descubrimiento muy importante sobre nosotros mismos y, por supuesto, es una base no solo para un simple sentimiento de 'sí, soy importante', sino que puede perfeccionarse una vez elaborado por la cultura y el lenguaje, puede darnos la sensación de que tenemos un alma inmortal.

En este punto, el lector podría pensar que es una idea extraña que un científico tome en serio, pero yo la tomo muy en serio. La creencia en el alma ha sido una de las fuerzas impulsoras de la historia humana y, de hecho, es responsable de la mayoría de las cosas significativas que los humanos han logrado alguna vez. Porque una vez que creemos en las almas y en su importancia, así como en las almas de otras personas y en su importancia, esto nos brinda nuevas ambiciones sobre lo que queremos dejar atrás y lo que queremos lograr en nuestras propias vidas.

Ahora, por supuesto, esto se trata de seres humanos. No creo que los perros, por ejemplo, tengan el sentido de ser almas. Pero ellos, sin embargo, se sienten como algún tipo de individuo significativo y, para ellos, debemos contar una historia un poco diferente. Para los humanos, la conciencia ha tenido resultados bastante inesperados y maravillosos. He descrito esto en mi libro más reciente, creo que es la joya de la corona de la evolución biológica y debemos tomar esta idea en serio.

Pregunta: El Profesor argumenta en su nuevo libro "Sentience: the Invention of Consciousness" que la sensibilidad está limitada a mamíferos y aves debido a una característica fisiológica muy específica: considera que son sensibles porque tienen sangre caliente. Esto es de hecho una característica curiosa e innovadora

que no se había considerado antes como una explicación plausible para la conciencia en su forma sensitiva. ¿Qué le llevó a esta idea?"

Nicholas Humphrey: Esta no es una tesis científica firme, pero creo que lo que debemos aceptar es que la sensibilidad, la conciencia fenomenal, es un desarrollo relativamente tardío en la evolución y no se remonta a los organismos primitivos. De muchas maneras, muchas personas, incluyendo a mi gran amigo y colega Daniel Dennett, piensan que los animales son sensibles desde el principio hasta el final, es un espectro entre organismos con más sensibilidad y otros con menos sensibilidad.

No creo que eso sea correcto, pienso que la sensibilidad es como un umbral, que surge en una etapa específica de la evolución porque implica un tipo particular de mecanismo cerebral y sus respectivos ciclos de retroalimentación. Y solo existe cuando es necesario. No servirá de nada para un animal que no se piensa a sí mismo, que no se relaciona con otros animales, que no tiene que ser, como ya he dicho, "psicólogos".

Cuando empecé mi investigación hace muchos años, comencé cuestionando "bueno, ¿Entonces cuándo surgió la conciencia?" y me di cuenta de que hay una transición en la evolución de los vertebrados que los filósofos y, de hecho, los biólogos, no tomaron muy en serio al respecto: cuando los animales se volvieron de sangre caliente.

Hace unos 200 millones de años, los dinosaurios comenzaron a tener sangre caliente y, posteriormente, las aves y mamíferos cercanos a ellos también la adquirieron. Esto significaba que su relación completa con el entorno había cambiado, se convirtieron en seres autónomos, independientes de su entorno físico inmediato, capaces de desplazarse donde quisieran, activos día y noche, entre otras cosas. Creo que esto les proporcionó un nuevo y fuerte sentido de individualidad, de autonomía, en ese momento, comenzaron a tener una utilidad para el concepto de un "yo".

Pero algo extraordinario que sucedió fue que, a medida que nuestros cerebros se calentaron, la velocidad de las células nerviosas también se incrementó. Aumentar la temperatura de doce o quince grados a treinta y siete o treinta y ocho grados en humanos y mamíferos, y cuarenta grados en aves, triplicó la velocidad eléctrica de las células nerviosas. Esto significó que nuestros cerebros y los de nuestros ancestros comenzaron a funcionar mucho más rápido que los de cualquier otro animal.

Creo que esto produjo una reorganización crucial en el cerebro. Permitted el desarrollo de ciertas formas de retroalimentación que simplemente no habrían ocurrido de otra manera. Por lo tanto, creo que hay una conexión tanto con un estilo de vida diferente como con la necesidad de una forma de pensarse a uno mismo,

porque ahora soy una criatura independiente del entorno, lo cual acompañó este cambio cerebral y permitió una nueva imagen psicológica de lo que es el cerebro.

Pregunta: El profesor está familiarizado con el trabajo de la filósofa Susan Schneider, quien ha propuesto una especie de prueba de conciencia para la inteligencia artificial, con el fin de determinar si una máquina artificial es consciente o no. ¿Cree usted que estas sugerencias avanzadas por la filósofa podrían proporcionar detalles específicos para crear tal prueba de conciencia artificial? En otras palabras, sabemos que la naturaleza ha creado la conciencia humana, pero ¿Cree usted que seremos capaces de crear o inventar conciencia artificial?

Nicholas Humphrey: Bueno, hay dos cuestiones diferentes aquí. Susan Schneider sugiere algunos tests que podríamos usar. Son tests muy buenos, bastante cercanos al tipo de tests que yo sugeriría, excepto que ella no toma tan en serio el lado social de la conciencia. Pero para mí, necesitamos esa capacidad, ya que permite entender la mente de otras criaturas, como nosotros. Creo que ella debería haber añadido ese factor a su lista de criterios para estar más cerca de un test diagnóstico, un test de conciencia en una máquina.

Sobre las máquinas, no hay dudas en este momento. Las personas han especulado si estos nuevos lenguajes como ChatGPT podrían ser conscientes. David Chalmers cree, increíblemente para mí, que hay un 10% de posibilidad de que ChatGPT ya sea consciente. Él es un filósofo: no creo que deba hacer afirmaciones tan contundentes. No se puede ser un 10% consciente, y no hay razón para pensar que cualquier cosa en este modelo de lenguaje sea consciente, ya que no tiene la necesidad ni los mecanismos para ello, y no muestra ninguno de los criterios diagnósticos de la consciencia.

Pero eso no significa que no podamos desarrollar una máquina que cumpla con estos criterios. Sin embargo, esto solo sucedería si diseñamos deliberadamente la máquina para tener esa capacidad, no ocurrirá por casualidad, sino porque creamos una máquina que procesa datos cada vez más rápido, o que es cada vez más inteligente.

La inteligencia no es lo mismo que la consciencia, por lo que no creo que las máquinas puedan alcanzarla hasta que podamos introducirla deliberadamente en ellas. Y, en este momento, está claro que no sabemos cómo hacerlo. Lo que defiendo en mi libro sugiere el tipo de cosas que necesitaríamos construir en una máquina para que se vuelva consciente.

Pero entonces tenemos que preguntarnos: ¿Por qué deberíamos hacer esto? Una razón podría ser, porque queremos, queremos que las máquinas tengan las

mismas capacidades de lectura de mentes y sensaciones, y adquieran una importancia psicológica —y así sucesivamente— como la que tenemos los seres humanos. Así, podríamos tomar una lección de la naturaleza y aplicarla a las máquinas, y tal vez desarrollar máquinas conscientes. Y quizás algún día solicitemos a nuestras máquinas conscientes que realicen nuestro trabajo mucho más allá de la tierra, en el espacio extraterrestre, por ejemplo.

Los humanos nunca serán capaces de vivir en una galaxia distante. Pero las máquinas podrían cumplir ese propósito, podríamos diseñarlas para que lleguen a otros lugares del universo. Sin embargo, si quieren establecer ese estilo de vida, si quieren estar lo suficientemente interesadas como para comenzar a desarrollar una ciencia de su nuevo entorno y considerar qué tipo de cultura quieren desarrollar, necesitarán tener confianza en su propia importancia. Entonces, tal vez primero necesitemos preguntarnos si nosotros, los humanos, queremos que esas máquinas tengan conciencia fenomenal o no.

Pregunta: El Profesor Humphrey argumenta también que algunos animales particulares son conscientes y otros no lo son en absoluto. Sabemos que la "conciencia" generalmente es una especie de propiedad moral que cualquier organismo necesita poseer para tener estatus moral y ser considerado un

agente de derechos. Siguiendo esto, ¿qué tipo de implicaciones éticas específicas cree que tiene su investigación, por ejemplo, en relación con la ética animal y los derechos de los animales en general?

Nicholas Humphrey: Ciertamente significa que, si está en lo cierto, debemos ser cautelosos antes de atribuir consciencia a otras criaturas. No debemos simplemente darles el beneficio de la duda y suponer que la consciencia estará presente solo porque tienen sistemas nerviosos complejos o muestran niveles elevados de inteligencia, que es lo que tiende a hacer la gente en este momento. La mayor parte de los escritos sobre consciencia y animales se centran en cuán inteligentes son, y no en cuán conscientes son en el sentido de tener una consciencia fenomenal. He discutido un poco este aspecto en el libro, no enfatizo mucho cuándo y por qué deberíamos considerar a otros animales con estatus moral solo porque tienen consciencia como la nuestra.

Mi opinión general es que, como científico y no necesariamente como ciudadano, debemos asumir que la mayoría de los animales no son conscientes, no creo que las langostas sean conscientes, no creo que los pulpos sean conscientes, y ciertamente no pienso que las lombrices lo sean. Esto significa que podemos reconsiderar los tipos de leyes que han sido aprobadas en todo el mundo sobre la consciencia animal. No sé cuál es la situación en Portugal, pero en Gran Bretaña

se aprobó una ley que establece que las langostas son conscientes. Como resultado, es ilegal hervir una langosta viva. Puede haber muchas razones para no querer hervir una langosta viva, pero ser consciente no es una de ellas, en mi opinión.

Pregunta: ¿Cuándo considera que la conciencia "aparece" en los seres humanos? ¿Cree que comienza a desarrollarse durante la gestación, quizás incluso en el útero materno, o prefiere establecer su inicio en la primera infancia? Esta cuestión puede tener profundas implicaciones para nuestra comprensión del desarrollo humano y los derechos éticos relacionados con la vida y el bienestar.

Nicholas Humphrey: Esta es una pregunta muy importante e interesante. Creo que no es durante la gestación o en el útero inicialmente, y una de las razones para afirmarlo es que los bebés humanos, al nacer, aún no tienen el cerebro mielinizado, las vainas de mielina no cubren las células porque aún no se han desarrollado, y ciertas áreas son disfuncionales debido a esto. Por ejemplo, el córtex visual no funciona en un recién nacido humano. Por lo tanto, creo que los niños humanos, si pueden ver —y sabemos que pueden ver por su comportamiento— deben tener algo similar a la visión ciega. Pero en ese punto, no tienen experiencia fenomenal.

Una vez más, soy parte de una minoría al afirmar esto. La mayoría de mis colegas no creen que sea posible ver sin tener sensaciones. Incluso si conocen casos de visión ciega, no los toman en serio en su práctica clínica diaria.

Si vemos a un bebé que claramente responde a la sonrisa de su madre, presumimos que el bebé está viendo el rostro de la madre de la misma manera que nosotros lo haríamos. No creo que esto sea cierto: pienso que alrededor de los tres o cuatro meses, cuando el cerebro se mieliniza, los circuitos serían suficientes para sostener una visión con experiencia fenomenal.

Pregunta: Para concluir, António Damásio, el neurocientífico portugués que ha publicado varias obras sobre estos temas, habla de diferentes tipos de “yo”, como el “yo autobiográfico”, que puede entenderse como la conciencia en cierto sentido. Pero también habla del “proto-yo” y del “yo nuclear”, como bases del “yo autobiográfico”. ¿Dónde se sitúa usted en este esquema conceptual sobre la personalidad?

Nicholas Humphrey: Creo que Antonio Damasio y yo estamos de acuerdo en muchas cosas, hemos discutido mucho sobre esto en diferentes contextos. Sin embargo, quiero ser mucho más específico que él al decir que la base del “yo” son las sensaciones y es sobre

eso que todo lo demás se construye. El sujeto de nuestros estados mentales es el "yo", construido sobre sensaciones. Y si seguimos desarrollándolo, es claro que existimos como Yos sociales, como Yos autobiográficos y así sucesivamente.

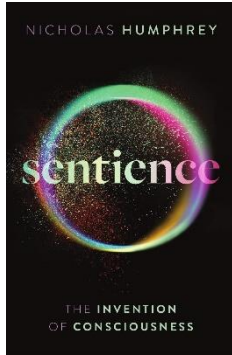
Aunque creo que él afirma sin mayores consecuencias que "los Yos existen en todos estos niveles diferentes, y cada uno de ellos es importante en aspectos diferentes", pienso que él evita la cuestión de comprender por qué alguien debería ser extremadamente importante para sí mismo en relación con los demás.

Mi creencia es que si el individuo pierde el "yo fenomenal", o el "yo central", no quedará nada. No puede mantener un "yo autobiográfico" o un "yo social" si pierde la fenomenología. Y lo interesante es que, en casi todos los casos de disociación y ausencia, donde los Yos de las personas se desintegran, estas permanecen realmente presentes teniendo Yoes fenomenales centrados en el dolor, los colores, las luces, y así sucesivamente.

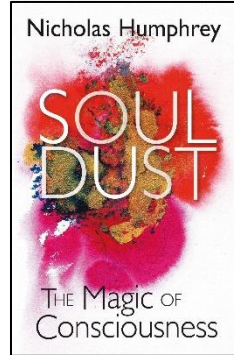
Pero no es eso lo que parece suceder, si fuera así, la persona no existiría. De hecho, hay casos en los que los pacientes afirman que no existen. En el Síndrome de Cotard, el paciente insiste en que ha muerto. Y cuando el médico dice "bueno, pero realmente no ha muerto, porque estoy hablando contigo ahora mismo, aquí en mi consulta", el paciente responde: "No, yo he muerto,

ya no estoy ahí". Creo que lo que están tratando de expresar es que el "yo fenomenal" ya no existe para ellos.

Libros de Nicholas Humphrey



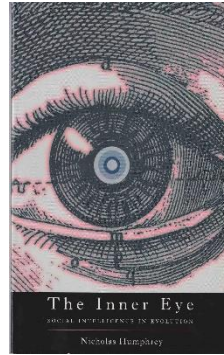
2023



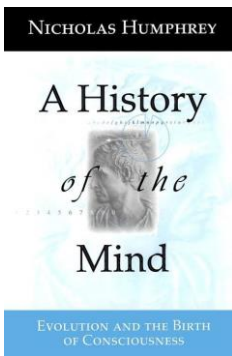
2011



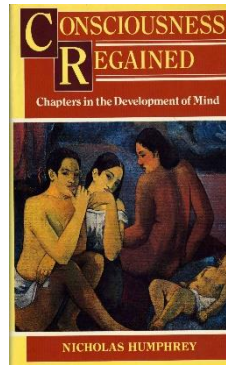
2006



2002



1998



1983

IX. Diálogo con Sir Roger Penrose



Sir Roger Penrose es profesor emérito de Matemáticas en la Universidad de Oxford y miembro emérito del Wadham College, Oxford, además de miembro honorario del St. John's College, Cambridge, y University College London.

Fue nombrado caballero por sus servicios a la ciencia en 1994 por la Reina Isabel II.

Em 1969, junto con Stephen Hawking, Penrose demostró que toda la materia dentro de un agujero negro colapsa hacia una singularidad.

Roger Penrose recibió varios premios y distinciones, incluyendo el premio Wolf de Física en 1988 y el premio Nobel de Física en 2020.

Él es autor de varios libros enfocados a la naturaleza de la conciencia. Como son: *The Emperor's New Mind* (1989), *Shadows of the Mind* (1994), *The Large, the Small and Human Mind* (1997), *The Road to Reality: A Complete Guide to the Laws of the Universe* (2007) o *Cycles of Time: An Extraordinary New View of the Universe* (2013).

Más información: maths.ox.ac.uk/people/roger.penrose

Pregunta: El profesor Penrose es uno de los científicos más relevantes del siglo XXI, habiendo ganado el Premio Nobel de Física en 2020 al demostrar que Albert Einstein estaba equivocado al negar la existencia de los agujeros negros. Su papel fue demostrar que, matemáticamente, la existencia de los agujeros negros es una consecuencia de la teoría de la relatividad de Einstein. Además de este interés, también ha investigado la naturaleza de la conciencia. ¿Cuándo comenzó a interesarse por este tema de la conciencia?

Roger Penrose: Puedo rastrear mi interés por la conciencia cuando era estudiante de doctorado, una de las principales influencias en la forma en que pienso sobre la conciencia ocurrió precisamente debido a una serie de conferencias a las que asistí en Cambridge mientras estaba estudiando (esto fue hace mucho tiempo, como puede imaginar).

Estaba estudiando Matemáticas Puras en ese momento, específicamente Geometría Algebraica, y tomé tres cursos que no tenían relación con ese tema. Uno de esos cursos se centraba en la relatividad general, impartido por Hermann Bondi, quien me influyó considerablemente. Luego, cursé una clase con el gran físico Paul Dirac sobre mecánica cuántica, lo cual también fue muy importante para mí. Y el tercer curso lo impartió un hombre llamado S.W.P Steen, especializado en lógica matemática y Máquinas de Turing, a partir del cual comprendí qué significaba la

idea de "computabilidad" en un sentido técnico o matemático.

Pero este profesor de Lógica también explicó en ese curso los dos teoremas principales de Gödel, y quedé absolutamente sorprendido por sus ideas sobre los teoremas, ya que cuando los escuché por primera vez, se me transmitió la idea de que había cosas en matemática que no podíamos demostrar matemáticamente. De hecho, no me gustó esa tesis desde el principio.

Lo que hice entonces fue inscribirme en este curso y me di cuenta de que, después de todo, no era exactamente así. Lo que ocurre realmente es lo siguiente: si usamos un cierto método de prueba que podría, en principio, ser implementado en una computadora, esta intentaría probar o refutar un resultado específico.

A partir de eso, podemos formular una proposición - una proposición matemática- sobre números, el tipo de cosa que el teorema pretende abordar y que puede, en términos generales, codificar la afirmación "no soy demostrable por estas reglas". Ahora, ¿es esta proposición verdadera o falsa? Supongamos que es falsa: entonces indicaría que es demostrable por esas reglas.

Así que lo que sigue es que debemos creer que cualquier cosa que pueda ser probada por las reglas es

verdadera, creo que ese es el punto central del teorema.

Elegimos las reglas de manera que cualquier cosa demostrable por esas reglas debe ser verdadera: si fuera falsa, entonces sería demostrable por las reglas y, por lo tanto, sería verdadera. Y si fuera verdadera, entonces es verdadera y no puede ser probada por las reglas, esa es la conclusión. Podemos ver que cierta afirmación es definitivamente verdadera y que no puede ser probada por las reglas. Pero, ¿Cómo podemos saber si realmente es "verdadera"?

Sabemos si es verdadera por la fe en las reglas, en la creencia de que las reglas realmente demuestran cosas. Si indican "sí, es verdad", entonces creemos en esas reglas, porque las comprendemos, de hecho, entendemos qué significan las reglas y no las seguimos solo por seguir.

Esa comprensión fue fundamental para mí. Entender lo que eso significa es más poderoso que simplemente seguir las reglas, ya que esta comprensión es algo consciente, que requiere nuestra conciencia.

Siguiendo estas ideas, argumenté que el pensamiento consciente trasciende simplemente "seguir reglas" y la conclusión a la que llegué es que no somos computadoras, alcancé esta visión de que el pensamiento consciente, sea cual sea, no es algorítmico porque implica algo más allá de seguir reglas.

Mi libro "The Emperor's New Mind" (1989) fue un intento de transmitir este punto y también de explicar diferentes temas sobre matemáticas y física, que me apasionan mucho, tratando de hacerlo no tan difícil para que las personas pudieran interesarse y aprender del libro.

Pensé que, al terminar el libro, habría aprendido lo suficiente sobre la neurofisiología humana para entender lo que posiblemente podría ser no computable en la acción del cerebro, pero no logré alcanzar ese logro. Llegué a la conclusión de que, aunque entendía los procedimientos, en ese momento creía que era necesario tener en cuenta el colapso de la función de onda en la mecánica cuántica, y esa es una parte muy relevante de todo mi argumento ¿Qué es lo que, en la física, puede trascender la computación?

Según la teoría dominante hasta ahora, tendríamos que encontrar algo en la física para explicar la conciencia, y los fisicalistas piensan que lo que sucede en nuestros cerebros es parte del mundo físico. Pero, ¿qué parte del mundo físico sería relevante para la conciencia? Mi argumento es que no se trata de la física "normal" que usamos en nuestra vida diaria, sino que tal vez sea algo que va más allá de esa física. Cuando escribí el libro, no tenía idea de qué podría haber en el cerebro que pudiera ir más allá de la física "normal", así que terminé el libro sin ninguna respuesta convincente sobre la conciencia.

Pregunta: ¿Podemos entonces afirmar que su tesis sobre la conciencia es una consecuencia directa del Teorema de Gödel? Dado que su afirmación es que la comprensión matemática es una característica de la conciencia y no es computacional, es decir, no es computable por principio o definición.

Roger Penrose: Sí, es exactamente eso.

Pregunta: Siguiendo esta tesis, ¿por qué cree que tantas personas involucradas en la Inteligencia Artificial creen que la mente es una especie de computadora digital y que podemos hacer ingeniería inversa del cerebro humano? ¿Por qué cree que tantos pensadores –bastante inteligentes– defienden estas creencias filosóficas sobre la mente y el cerebro?

Roger Penrose: Creo que básicamente estos pensadores no han seguido o no han comprendido mi argumento. Existen varias críticas que pueden señalarse a mi argumento, y la más fuerte es la siguiente: si no sabemos qué tipo de algoritmos usa nuestro cerebro, entonces no podemos construir los resultados de Gödel y, por lo tanto, el argumento no funciona.

El contraargumento que presento contra esta crítica en mi otro libro, "Shadows of the Mind", es que, si suponemos que existe un algoritmo en nuestras cabezas (algo que no sabemos si es verdad), ¿Cómo

surgió ese algoritmo en primer lugar? Debemos suponer que vino, como todo lo demás, de la selección natural, tiene que ser algo que nos ofrece una ventaja selectiva.

La imagen metafórica que usé en el libro es imaginar a nuestros antepasados haciendo cosas útiles, como construir refugios, domesticar animales, plantar... y en primer plano, vemos a alguien teniendo la idea de cómo construir una trampa para mamuts: todas esas personas están haciendo algo que les proporciona ventajas selectivas. Pero en otro plano de la imagen, encontramos a este pobre matemático que está trabajando en algún teorema y está a punto de ser devorado por un tigre dientes de sable. La moraleja de esta metáfora es que hacer este tipo de matemáticas sofisticadas no trae ninguna ventaja selectiva, aunque pueda producir varios resultados matemáticos.

Uno de esos resultados matemáticos sofisticados que más me gusta se llama el Teorema de Goodstein. Es un resultado maravilloso que puede explicarse a personas que no poseen un conocimiento muy profundo de matemáticas. El lector solo necesita saber qué es elevar un número a una potencia en notación matemática.

El teorema de Goodstein afirma que, si sigues determinados procedimientos "A" y "B" repetidamente, siendo que cada uno de esos procedimientos es muy simple, esto no dura para siempre y encuentra un fin.

El problema es que esto requiere tantas etapas que es completamente ridículo intentar demostrar, de hecho, el teorema, lo que significa que no puede tener ninguna ventaja selectiva. Sin embargo, ¿cómo sabemos que este teorema es verdadero?

Bueno, sabemos que es verdad gracias a algunos resultados maravillosos del matemático Georg Cantor, y lo más sorprendente es que no necesitamos ser Cantor para entender esos resultados.

Al contrario, simplemente entender el razonamiento de Cantor no requiere mucho de cada uno de nosotros, implica un ligero cambio en el punto de vista habitual, pero el razonamiento en sí no es tan difícil de comprender. ¿Pero cómo sabemos esto? ¿Cómo podríamos haber evolucionado si fuéramos solo algoritmos? No existe ningún algoritmo con esa sofisticación.

Esto significa que la calidad general de la comprensión, sea cual sea esa calidad (sin pretender definirla exactamente), debería ser una característica de la conciencia. Por supuesto, no estoy afirmando que esto agote la conciencia, ya que esta abarca muchos otros aspectos además de la comprensión matemática, como la percepción del color azul, por ejemplo, o la sensación de dolor. Hay muchos aspectos de la conciencia que no tienen nada que ver directamente con las matemáticas.

Pero mi argumento es que, si podemos entender estos resultados en matemáticas, deben estar más allá de la computación. No podemos decir que el Teorema de Goodstein esté fuera del alcance de la computación, porque una vez que conozcamos la entrada de Cantor, podríamos poner una versión específica de ello en una computadora. Sin embargo, desde el punto de vista de la selección natural, esto no es creíble en absoluto, no ofrece ninguna ventaja selectiva, ya que está muy alejado del uso normal de nuestros cerebros.

Lo que estoy intentando defender con mi argumento es que la calidad general de comprensión de las cosas no es algorítmica. No estoy hablando aquí de cómo es experimentar el color azul; mi tesis central es que, sea lo que sea la conciencia, es algo que va más allá de un procedimiento computacional.

Pregunta: Una de las opiniones más comunes en la neurociencia actual sobre la conciencia esencialmente sostiene que esta cualidad subjetiva es una especie de propiedad emergente del cerebro, resultado de la actividad neuronal y del intercambio de información entre las neuronas. Sin embargo, el profesor Penrose no está de acuerdo con esta visión, ¿verdad?

Roger Penrose: Lo que pienso es que no creo que haya alguna "otra cosa" misteriosa que cree la conciencia en nuestras cabezas. Creo que tiene que ver

fundamentalmente con la neurofisiología que ocurre dentro de la cabeza: simplemente aún no sabemos exactamente cómo se crea la experiencia subjetiva en el cerebro.

Pregunta: Pero en cuanto a la naturaleza de la conciencia, ¿El profesor cree que la neurociencia puede resolver este problema por sí sola, o cree que necesitamos lo que el profesor Penrose llama "una nueva física" para poder enmarcar este fenómeno tan particular dentro de la imagen científica del mundo?

Roger Penrose: Sí, creo que necesitamos una "nueva física", aunque mi argumento tiene varias partes, por ejemplo: "¿Cuáles son las otras partes de la física que conocemos?"

Podemos centrarnos en una de las principales teorías que tenemos, la física newtoniana y ponerla en un ordenador, como algunos investigadores hacen. ¿Y la teoría general de la relatividad de Einstein? Los científicos han descubierto cómo los agujeros negros espiral entre sí y qué tipo de señal emana de esa interacción; claramente, este es un procedimiento computacional. Necesitamos determinar qué sistemas se pueden poner en un ordenador. Yo diría que no es la relatividad general.

¿Y la mecánica cuántica? Bueno, en la mecánica cuántica existe la ecuación de Schrödinger que también

puede ser implementada en un computador. Volvamos a uno de los cursos que tomé cuando era estudiante de doctorado de Paul Dirac. En su primera clase, él habló sobre el principio de superposición: la idea de que un átomo puede estar "aquí" o "allá". En la mecánica cuántica encontramos estados donde el átomo está "aquí" y "allá" al mismo tiempo; este es el principio fundamental de la mecánica cuántica.

Siguiendo esto, el profesor Dirac tomó un trozo de tiza y lo rompió en dos partes, intentando describir cómo ambos pedazos de tiza podían estar "aquí" y "allá" al mismo tiempo. Lo más importante de este episodio es que yo estaba presente en la clase, pero mi mente estaba divagando: estaba mirando por la ventana, pensando en algo completamente diferente. Cuando mi mente volvió a la clase, él ya había cambiado de tema. Recuerdo que estaba hablando sobre una energía asociada al trozo de tiza, pero no podía entender qué tenía que ver eso con la superposición.

Entonces, no pude escuchar la respuesta, lo cual resultó ser, al final, algo positivo. Creo que probablemente él estaba tratando de tranquilizarnos de alguna manera, para que no nos preocupáramos por este problema, porque las energías involucradas son muy altas y no son relevantes a este nivel de descripción. Pero me quedé con la sensación de que necesitábamos algo nuevo, necesitábamos una "nueva física". Y creo que es precisamente en este problema donde necesitamos

una nueva física, ya que la ecuación de Schrödinger no describe realmente el mundo, como todos los físicos saben, pero fingen olvidarlo. Ellos saben esto, pero no lo afirman en voz alta.

Schrödinger estaba muy consciente de esto, ya que describió a las personas después de saber que había puesto a ese pobre gato en la caja, colocándolo en un estado de estar "vivo" y "muerto" al mismo tiempo, en un estado de superposición. Sin embargo, en mi opinión, las personas no entendieron lo que Schrödinger intentaba decir con esta experiencia mental.

Los físicos suelen decir algo como "bueno, si tuviéramos una experiencia suficientemente sofisticada, podríamos hacer que un gato esté vivo y muerto al mismo tiempo, pero aún no hemos llegado allí". Pero esto no es lo que Schrödinger estaba tratando de afirmar. Lo que él estaba intentando decir era que tal escenario es simplemente ridículo, dado que no podemos tener un gato vivo y muerto al mismo tiempo, lo cual indica que hay algo seriamente incorrecto con su propia ecuación (la de Schrödinger). De hecho, él estaba intentando argumentar en contra de su propia ecuación, estaba afirmando que la ecuación de Schrödinger no explica cómo funciona el mundo en su totalidad.

Pregunta: ¿Podrías ampliar la idea de que podría haber un hiato o brecha en la mecánica cuántica, sugiriendo que falta algo para considerarla una teoría completa? ¿De qué manera visualizas esta incompletitud y cómo podría relacionarse con nuestra comprensión de la conciencia?

Roger Penrose: Sí, creo que hay una brecha considerable. Pero es una especie de brecha separada, ya que este es un argumento dentro del ámbito de la física. Lo que estoy diciendo es que hay una brecha en la mecánica cuántica y tal vez sea ahí donde reside esa no-computabilidad que mencioné anteriormente.

Cuando escribí "The Emperor's New Mind", ese era el tono que intentaba transmitir. El punto débil del libro era que no tenía idea de cómo los neuronas podían lograr eso. Pensé que al final del libro habría aprendido lo suficiente sobre neurofisiología para ver dónde podría haber una brecha de ese tipo, pero no lo logré. Así que medio que abandoné el libro, pero tuve que terminarlo de alguna manera y no estaba muy contento con todo lo que añadí en ese momento.

Fue entonces cuando Stuart Hammeroff, quien había leído mi libro, decidió escribirme una carta en la que especificaba ciertas estructuras anatómicas que yo desconocía, conocidas como los microtúbulos. Él tenía una teoría que comenzó a recibir apoyo desde otros ángulos de investigación, como por ejemplo, cómo los anestésicos pueden crear diferentes estados de

conciencia. El enfoque de Hammeroff era el siguiente: Existe algo que puede desconectar la conciencia de manera reversible: podemos apagarla y encenderla de nuevo. ¿Cómo? Por ejemplo, podemos ser inducidos a un estado de coma y perder la conciencia, y luego despertar con la conciencia volviendo.

Siendo anestesiólogo, lo que Hammeroff está intentando hacer no es simplemente dejar a los pacientes inconscientes y luego despertarlos de nuevo, sino que está tratando de entender qué es lo que realmente sucede con esas sustancias que "desconectan" la conciencia. Esto es muy interesante porque la conclusión no es un proceso químico, lo que une estas diferentes sustancias está relacionado con un proceso físico, no químico.

Y la pregunta que surge aquí es: ¿Qué está realmente ocurriendo? Lo interesante es que, aunque Hammeroff y yo nos conocimos personalmente en la década de 1990, en aquel entonces este enfoque era considerado muy controvertido. Sin embargo, está emergiendo como una de las perspectivas que los científicos y filósofos toman en serio, aparte de los puntos de vista típicos que normalmente se consideran sobre la conciencia, como el dualismo o el reduccionismo.

Para mí, eso es bastante reconfortante. Saber que las personas están tomando en serio nuestra teoría sobre la conciencia, con experiencias muy interesantes sobre la acción general de los anestésicos en la conciencia y

qué estructuras específicas del cerebro son afectadas por esas sustancias. Parece que estas estructuras -los microtúbulos- también pueden ser un ingrediente importante en toda esta historia de la conciencia. Las cosas están avanzando hacia una dirección en la que realmente podremos confirmar si tenemos una teoría correcta de la conciencia o no.

Pregunta: Vamos a profundizar un poco más en su modelo teórico de la conciencia. Básicamente, el profesor es crítico de la llamada "interpretación de Copenhague" de la mecánica cuántica, argumentando que lo que sucede es contrario a lo que esta interpretación describe. Tenemos el colapso de la función de onda, pero no necesitamos ningún sujeto consciente para que esto ocurra, no necesitamos un observador. Siguiendo esto, ¿Cómo puede surgir la conciencia de este proceso de colapso de la función de onda a través de la influencia de las fuerzas gravitacionales, que es lo que el profesor defiende?

Roger Penrose: Sí, como dijo Steven, es lo contrario. Hay varios elementos que propongo. Uno que no mencioné es que el colapso de la función de onda, en mi opinión, tiene que ver con la combinación de la gravedad y la mecánica cuántica. Este punto particular aún no ha sido confirmado ni refutado experimentalmente, pero hay algunos experimentos

destinados a abordar este problema. También hay razones teóricas para defender este punto de vista.

Específicamente, tengo un argumento que presenté hace algunas décadas que busca mostrar que existe una incompatibilidad entre los principios fundamentales de la mecánica cuántica, que involucra el principio de superposición que mencioné, y el principio de la relatividad general, que es el principio de Galileo, que básicamente afirma que podemos cancelar la gravedad en caída libre.

Galileo imaginó que si una piedra grande y una pequeña caían desde la torre inclinada de Pisa sin atmósfera, caerían juntas a la misma velocidad, y si hubiera un insecto sentado en una de las piedras observando a la otra durante la caída, ese insecto pensaría que no hay gravedad. Este experimento mental muestra que podemos cancelar la gravedad en caída libre, y este es el principio de equivalencia, que es la piedra angular de la relatividad general de Einstein. Sin embargo, este principio fundamental es inconsistente con los principios fundamentales de la mecánica cuántica, especialmente con el principio de superposición.

Mi argumento es que hay un conflicto ahí, y para resolver ese conflicto será necesaria una nueva teoría. Aunque no sé cuál es esa nueva teoría, lo que puedo decir es una estimación de cuándo esa teoría entra en funcionamiento y en qué tipo de escala podríamos

empezar a observar el colapso de la función de onda siendo un proceso físico que podemos realmente medir.

Pero como dijo Steven, es lo contrario de lo que muchos pensaban en los comienzos de la mecánica cuántica, es decir, que el colapso de la función de onda, que ocurre cuando hacemos una medición, depende del observador, de un ser consciente. Personas como John von Neumann y Eugene Wigner defendieron este tipo de visión. De hecho, conversé personalmente con Wigner sobre esto en una ocasión y descubrí que no era tan dogmático como algunas personas piensan: ciertamente estaba abierto a otras alternativas.

Pero como dijo Steven, es lo contrario de lo que muchos pensaban en los comienzos de la mecánica cuántica, es decir, que el colapso de la función de onda, que ocurre cuando hacemos una medición, depende del observador, de un ser consciente. Personas como John von Neumann y Eugene Wigner defendieron este tipo de visión. De hecho, conversé personalmente con Wigner sobre esto en una ocasión y descubrí que no era tan dogmático como algunas personas piensan: ciertamente estaba abierto a otras alternativas.

Wigner parecía pensar que este era un punto de vista que debía ser tomado en serio, pero solo hasta cierto punto, porque podemos notar que realmente hay una inconsistencia y que no podemos confiar en seres conscientes para colapsar la función de onda. Creo que

tiene que ser un proceso físico, pero como dije, creo que es el cerebro el que, haciendo uso de ese proceso físico, produce la conciencia.

Pregunta: ¿Qué tan cerca crees que estamos de establecer una conexión genuina entre los estados de conciencia fenomenal, como emociones o sentimientos, y la reducción fisiológica de estados de superposición cuántica en el cerebro, particularmente dentro de los microtúbulos?

Roger Penrose: Claro que queda mucho camino por recorrer. No veo eso directamente. Diría, sin embargo, que estoy bastante impresionado por lo lejos que han llegado las cosas; no esperaba ni siquiera el progreso que se ha logrado. Recientemente estuve en una conferencia en Canadá donde hubo muchas charlas interesantes sobre estas ideas de manera seria y con experimentos. Me impresionó lo avanzado que está, pero, no obstante, no ha avanzado mucho en las direcciones de la conciencia real.

Los experimentos en microtúbulos, proteínas de tubulina o en la propagación nerviosa y cosas así, están comenzando a entenderse cosas que antes no se conocían, y creo que hay cosas muy interesantes e inesperadas por aprender de estos experimentos.

Más específicamente, sé que esto no forma parte de tu pregunta, pero hay algunos enigmas sobre cómo las

personas pueden reaccionar tan rápidamente, especialmente en juegos como el ping-pong o el tenis, pero especialmente el ping-pong donde tienes que reaccionar tan rápidamente. Solía jugar ese juego yo mismo y se puede ver que las decisiones conscientes podrían influir en lo que haces mucho más rápidamente de lo que deberían, porque cuando se analizan dónde están estas señales nerviosas y cuánto tiempo tardan y qué parte del cerebro debe estar involucrada, parece que no hay manera de que puedan actuar tan rápidamente, por lo que plantea muchos problemas muy curiosos.

Muchas personas dicen que esas cosas que crees que sabes que estás haciendo se hacen todas inconscientemente, así que en realidad no te dicen nada sobre la conciencia. Yo no creo en esto, ya que cuando solía jugar al ping-pong, pensaba que realmente estaba decidiendo si golpear la pelota hacia abajo de una manera u otra, y tomaba esa decisión muy rápidamente.

La reducción del estado cuántico es un fenómeno muy peculiar cuando intentas integrarlo con otros principios de la relatividad general y la mecánica cuántica, y al intentar combinar estas cosas, surgen relaciones muy extrañas en cuanto a la forma en que progresa el tiempo. No son inconsistentes, pero son muy peculiares, y nos indica que el tiempo en el que

pensamos que percibimos que las cosas están ocurriendo no es del todo como pensamos.

Hubo algunos experimentos iniciales, que en realidad describí en "La mente nueva del emperador", realizados por Benjamin Libet, donde tenía a un paciente en una operación cerebral (no entraré en detalles), pero era muy difícil explicar la línea temporal, qué sucedió antes y qué sucedió después: no tenía sentido alguno.

El argumento es que hay algo muy peculiar ocurriendo en relación con la naturaleza temporal de la experiencia, cuando crees que experimentas algo y cuando realmente experimentas algo, y para dar sentido a esto implica algunas cosas muy curiosas, las cuales, creo, serán muy intrigantes en cualquier teoría a la que lleguemos más adelante, pero estamos aún lejos de eso.

Pregunta: Para concluir y observando el fenómeno de la espiritualidad, hay muchos libros basados en temas como la sincronicidad cuántica, los chakras, el chi, entre muchos otros. Si el Profesor pudiera, muy brevemente, decirnos qué piensa sobre estos enfoques y esos libros que son éxitos de venta... ¿Cree que hay algo a lo que se pueda dar validez?

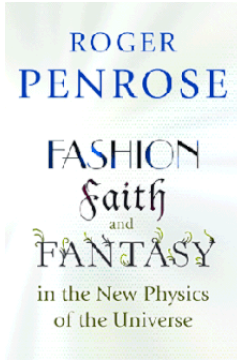
Roger Penrose: No soy seguidor de ninguna de estas cosas, debo confesar que no he leído mucho en esa

dirección, mi tendencia es no creer que haya mucho en ello. Podemos aprender algo sobre la conciencia al observar estas ideas, pero yo me he concentrado únicamente en un aspecto muy limitado de la conciencia, que es esta cualidad de comprensión, específicamente cuando se trata de ideas matemáticas.

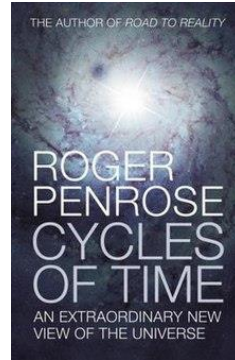
Me he alejado de esas ideas principalmente porque no veo cómo podemos hablar de ellas de manera coherente y precisa. Creo que ese es el problema. Soy matemático: me gusta poder integrar las cosas en una teoría donde pueda ver las consecuencias surgir, determinadas con precisión de manera clara por esa misma teoría.

No veo esas cosas de la manera que mencionaste, por lo tanto, no pierdo mi tiempo estudiándolas. No veo ningún beneficio, al menos para mí, aunque podría haber beneficio para otras personas.

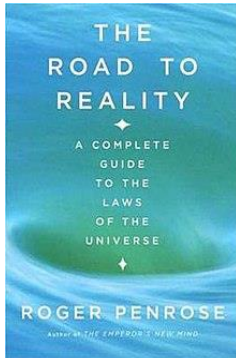
Libros de Sir Roger Penrose



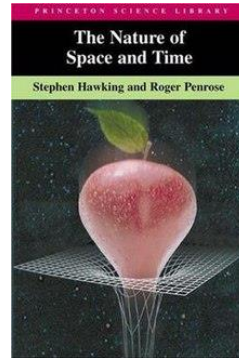
2016



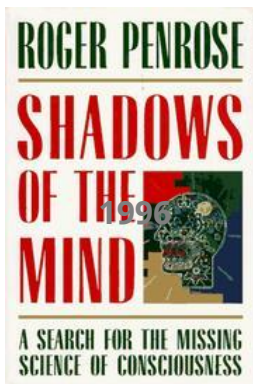
2010



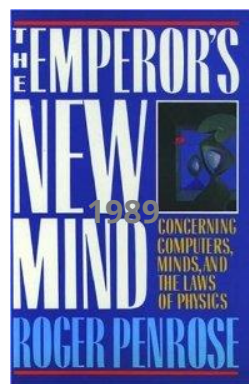
2004



1996



1996



1989

PARTE II

Cerebro

PARTE II

Cerebro

I. Introducción: Cerebro

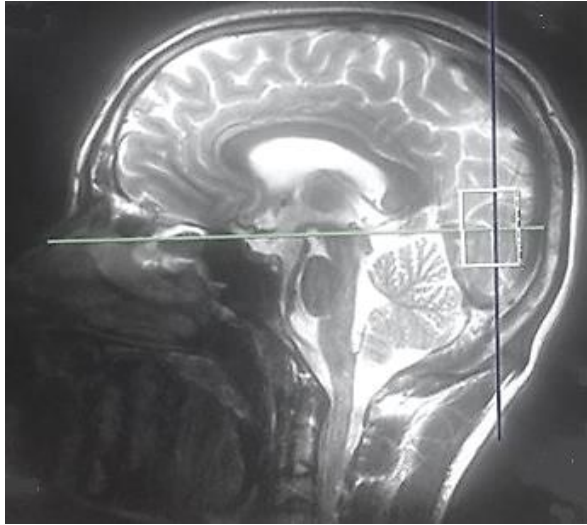


Imagen 13: fMRI de un cerebro.

La naturaleza filosófica y científica del cerebro, como el cerebro en la imagen anterior, ha sido objeto de debate durante cientos de años en diversas disciplinas. Actualmente conocemos mucho más sobre el cerebro de lo que nuestros abuelos conocían, por ejemplo, sabemos que en promedio pesa 1,5 kg; sabemos que está compuesto por varios tipos diferentes de neuronas que intercambian información electroquímica entre sí; y sabemos que estas neuronas

pueden formar diferentes redes neuronales relevantes para diversas funciones mentales, como la memoria a largo plazo, las emociones o la creatividad.

El cerebro humano tiene alrededor de 86 mil millones de neuronas conectadas a través de trillones de sinapsis y está dividido en varias regiones distintas. El núcleo más primitivo, similar a las antiguas estructuras de la evolución, se encarga de los instintos básicos de supervivencia. Justo después, encontramos el sistema límbico, responsable de las emociones y la formación de la memoria. Finalmente, tenemos la superficie del cerebro, llamada córtex cerebral, que está dividida en varios lóbulos y es responsable de los procesos cognitivos complejos.

A pesar de que actualmente conocemos mucho más sobre el cerebro -aunque aún sabemos muy poco, como veremos en esta segunda parte del libro- este órgano ha tenido una historia controversial desde que la humanidad comenzó a pensar filosóficamente sobre sí misma como un ser constituido de carne y hueso. En particular, dos grandes pensadores hicieron algunas consideraciones sobre la conexión del cerebro con la mente humana.

En un debate apasionado, Aristóteles de Estagira defendió, en el libro III de su obra "Las partes de los animals",⁵⁰ que el corazón era la sede y la fuente de las

⁵⁰ Versión en línea: penelope.uchicago.edu/aristotle/parts3.html

sensaciones, la conciencia, el intelecto y todo lo relevante en el ser humano, creyendo que el cerebro tenía una única función: enfriar la temperatura de la sangre. En contra de esta perspectiva, Hipócrates de Cós defendía en su libro "La enfermedad sagrada"⁵¹ que la perspectiva aristotélica estaba equivocada. El cerebro debería ser considerado como la fuente del placer, la alegría, la tristeza o el dolor, en lugar del corazón.

Aunque la antigua Grecia sea un punto de referencia histórico fascinante, se especula que las primeras consideraciones conocidas sobre el cerebro humano no comenzaron con los antiguos griegos, como comúnmente se asume, sino antes.

Al final de la Edad de Piedra, un período que comenzó hace aproximadamente 50,000 años en Europa y avanzó hacia el período Neolítico, una era que abarca desde el 3,000 hasta el 2,000 a.C., la humanidad desarrolló prácticas agrícolas inteligentes y domesticó ciertos animales.

En este período también surgieron la invención de herramientas de piedra pulida, la creación de cerámica y la formación de pequeños poblados, lo cual permitió que algo absolutamente fascinante pudiera ocurrir.

Del mismo período, se han encontrado registros fósiles de las primeras perforaciones del cráneo humano

⁵¹ Versión en línea: classics.mit.edu/Hippocrates/sacred.html

realizadas mediante procedimientos quirúrgicos con herramientas especializadas de piedra, lo que indica que el hombre neolítico utilizaba varios tipos de instrumentos afilados para llevar a cabo una operación conocida como "trepanación" (del griego "trypanon", que significa "perforar"), que es la primera práctica quirúrgica conocida realizada por la humanidad.⁵²

El primer cráneo reconocido como perforado fue descubierto accidentalmente en 1864 debido a... heces de pájaro. ¿Cuál sería la conexión entre cráneos y heces? En 1861, estalló la Guerra Civil estadounidense y se volvió necesario asegurar la fertilización para la producción agrícola, convirtiéndose en una de las prioridades de Abraham Lincoln.

Curiosamente, los mejores fertilizantes del mundo en ese momento provenían de América del Sur, obtenidos de una sustancia llamada "guano", que se extrae de las heces de varios animales y es rica en fósforo y nitrógeno. Por lo tanto, Ephraim G. Squier, un arqueólogo y periodista, fue enviado por Lincoln a Perú para asegurar la producción y el envío del fertilizante a los Estados Unidos.

Después de completar su tarea, Squier decidió explorar el país y llegó al cementerio inca en el Valle de Yucay,

⁵² Basado en: Weber, J. e Wahl, J. (2006) "Neurosurgical aspects of trepanations from Neolithic times", *International Journal of Osteoarchaeology*, 16: 536-545.

donde encontró un cráneo perforado que tenía un agujero rectangular de 15x17 mm en su estructura, algo que lo dejó asombrado, dado que no es común encontrar ángulos rectos en la naturaleza. Con este descubrimiento, Squier decidió regresar a casa y presentó el artefacto por primera vez ante la Academia de Medicina de Nueva York en 1865.

Desafortunadamente, no fue tomado en serio, ya que los científicos de esa organización no creían posible un conocimiento quirúrgico tan avanzado entre los pueblos indígenas, lo cual reflejaba una clara postura racista.

A pesar de eso, Squier no renunció a analizar el cráneo y decidió enviarlo a París, donde vivía quien era considerado en ese momento la máxima autoridad científica en el estudio del cerebro, el famoso fundador de la primera "Sociedad Antropológica de Francia", el médico y profesor Paul Broca.

Broca quedó completamente sorprendido al analizar detalladamente el cráneo y darse cuenta de que la perforación contenida en él claramente había sido realizada por una herramienta de corte especializada y que involucraba algún tipo de proto-cirugía, algo que hasta entonces se consideraba totalmente inviable.

Este fue el primero de muchos cráneos perforados que fueron descubiertos en las décadas siguientes, la mayoría de los cuales revelaron un detalle aún más

asombroso: evidencia de crecimiento óseo alrededor de las perforaciones. Ahora, ¿qué significa esto? Significa lo siguiente: que los individuos sometidos a estas cirugías primitivas pudieron haber sobrevivido varios meses o incluso años, lo cual es extraordinario.

Ahora bien, para asegurar un paso exitoso al "otro mundo", los egipcios creían que, para garantizar que el alma llegara a buen puerto, el cuerpo también debía ser preservado. Es esta creencia la que explica todo el proceso de momificación que se hizo famoso en varias obras literarias y cinematográficas.

Otra fuente histórica se encuentra en la región del Nilo, en el antiguo Egipto. Las grandes pirámides y los diversos procedimientos funerarios eran indicativos de que la vida después de la muerte tenía una gran relevancia para los antiguos egipcios.

Entonces, para asegurar un paso exitoso al "otro mundo", los egipcios creían que para garantizar que el alma llegara de manera segura, el cuerpo también debía ser preservado. Es esta creencia la que explica todo el proceso de momificación que se hizo famoso en varias obras literarias y cinematográficas.

De manera interesante, la parte más importante del cuerpo era el corazón, ya que se creía que representaba el "yo" de la persona. Los intestinos, los pulmones, el hígado y el estómago también se consideraban

importantes, siendo embalsamados y almacenados en vasijas canónicas junto a la momia.

En el lado opuesto de relevancia estaba el... cerebro, que, imagínese, simplemente se descartaba y se extraía a través de las fosas nasales utilizando un gancho de hierro. Según nuestro entendimiento, los antiguos egipcios atribuían poca importancia al cerebro en su relación con la mente o la conciencia.

A través de varios escritos egipcios -particularmente aquellos que adornan las paredes de las tumbas- sabemos, no obstante, que el corazón no solo era considerado como el depósito de las acciones terrenales del alma, sino que también poseía capacidades cognitivas y conativas (emociones y sentimientos).

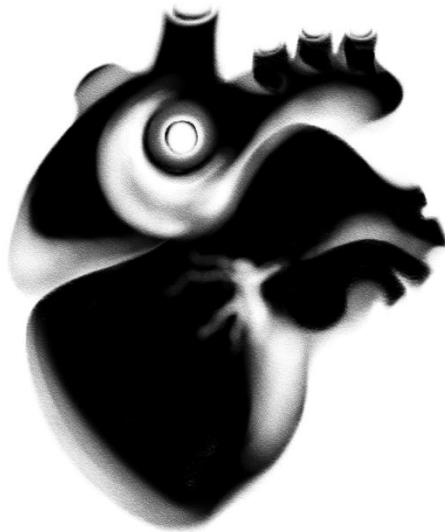


Imagen 14: "Pleasure" © Gaia Amoroso

Esta primacía del corazón fue mantenida, curiosamente, hasta los tiempos bíblicos. De hecho, cuando el "Antiguo Testamento" fue traducido al griego, se aceptó que el intelecto y las emociones del hombre residían en el corazón y no en el cerebro y no hay ni una sola referencia al cerebro en la Biblia.⁵³

No obstante, a pesar de este menosprecio hacia el cerebro, posiblemente podemos encontrar en los antiguos egipcios la referencia escrita más antigua al cerebro. Un papiro médico conocido como el "Papiro de Edwin Smith", que describe 48 tipos de lesiones en la cabeza y el cuello, junto con consejos sobre tratamiento e intervención quirúrgica y que contiene referencias a textos escritos hasta el 3.000 a.C.

Los autores del papiro parecen tener cierta comprensión de la función del cerebro, ya que describieron ciertas lesiones en la cabeza y las correlacionaron con varios síntomas como parálisis o pérdida del habla. Sin embargo, también cometieron errores graves, como el hecho de que una lesión en el hemisferio derecho causaría lesiones en el lado derecho del cuerpo, cuando sabemos que en realidad debería afectar al lado izquierdo.

⁵³ Basado en: York, G.K. e Steinberg, D.A. (2010) "Neurology in Ancient Egypt" In Finger, S., Boller, F. e Tyler, L. (eds) *Handbook of Clinical Neurology*, vol. 95, Elsevier: Amsterdam.

Otra referencia histórica fascinante sobre el cerebro se puede encontrar en las aventuras de Aquiles y Ulises, es decir, en los textos homéricos de la *Ilíada* y la *Odisea*, que inspiran este libro. Para los griegos, el "yo" estaba compuesto por diversas fuerzas, una de ellas siendo la psique, la fuerza vital que permitía mantener viva a la persona. Homero eventualmente identificaría lo mental en varias otras "formas" de alma que residían en el pecho.

La forma más importante de estas sería el "thymos", mencionado más de 450 veces en la *Ilíada*, que estaría ubicado en el diafragma, siendo la fuente de emociones que llevan a alguien a actuar. En cuanto a las capacidades intelectuales, estas formarían parte del "noos", que estaría localizado en el pecho. En la misma línea que los egipcios, los textos homéricos no atribuyen gran importancia al cerebro.

Hay, no obstante, una influencia lingüística muy importante de Homero, que introdujo en el vocabulario tres conceptos fundamentales: (1) "enkephalos", el propio cerebro; (2) "muelos", refiriéndose a la columna vertebral (de donde obtenemos la palabra 'médula'); (3) "sinew", que originó el concepto de "neurona".⁵⁴

A pesar de esta desconsideración del cerebro, podemos encontrar en la antigua Grecia algunos pensadores que,

⁵⁴ Basado en: Singer, C. (1957) *A Short History of Anatomy and Physiology from the Greeks to Harvey*, Dover: New York.

contrariamente a esta creencia compartida por los egipcios, defendieron que este órgano particular debería tener mayor relevancia que la que se le estaba dando hasta entonces.

Uno de estos pensadores era Alcmeón de Crotona, quien consideraba el cerebro como el órgano de las sensaciones, aunque todos sus textos se han perdido y no ha llegado ninguna referencia directa a nuestros días.

A pesar de eso, nos han llegado varias referencias indirectas que apuntan al brillante protoanatómico de este filósofo. Por ejemplo, Teofrasto de Eresos afirmaba que Alcmeón habría sido el primero en estudiar la anatomía de varios animales mediante técnicas de disección de los mismos, habiendo descubierto que existían dos canales que conectarían físicamente la parte posterior del ojo al cerebro, lo que conocemos actualmente como los nervios ópticos, lo que le llevó a defender que los sentidos estarían todos conectados al cerebro.

Otra referencia indirecta se puede encontrar en Aecio, contemporáneo de Alcmeón, quien defendía que este último era conocido por sostener que la inteligencia, al igual que las sensaciones, estarían relacionadas con el cerebro humano. Debido a estos y otros descubrimientos, el impacto de las investigaciones de

Alcmeón es comparable a los descubrimientos de Copérnico y Darwin.⁵⁵

Como vimos anteriormente, la otra gran referencia griega sobre la relevancia del cerebro es el fundador y padre de la medicina, Hipócrates. En el "Corpus Hippocraticum", hay múltiples referencias al cerebro, siendo el texto más relevante "De la enfermedad sagrada" que se enfoca en gran parte en un intento de comprender la naturaleza de la epilepsia.

Hipócrates es el primero en intentar demostrar que esta enfermedad no podría explicarse por la teoría médica vigente hasta entonces, la llamada "Teoría de la Posesión Demoníaca", donde se defendía que la epilepsia estaba ligada a una forma de posesión por un demonio particular como castigo divino por algún pecado cometido por esa persona.

El fundador de la medicina intentó refutar esta teoría, argumentando que era una noción defendida por charlatanes y curanderos que no buscaban realmente un tratamiento para esta enfermedad. Defendía que la epilepsia era una enfermedad del cerebro causada por un exceso de flema que bloqueaba el flujo de aire en los vasos sanguíneos. Ahora bien, dicho bloqueo solo podía corregirse a través de una convulsión epiléptica, cuya finalidad sería desbloquear ese bloqueo.

⁵⁵ Basado en: Mithen, S. (1996) *The Prehistory of the Mind*, Phoenix Books: Guernsey.

Además de esta explicación naturalista (que no apela a demonios oscuros), Hipócrates argumentó que el cerebro es responsable de toda nuestra actividad mental. Vale la pena prestar atención a las palabras del filósofo y médico:

“Debe ser de conocimiento general que la fuente de nuestro placer, alegría, risa y diversión, así como de nuestra tristeza, dolor, ansiedad y lágrimas, no es otra que el cerebro. Es este órgano el que nos permite pensar, ver y oír, y distinguir lo feo de lo bello, lo malo de lo bueno, lo agradable de lo desagradable... es también el cerebro el que es la sede de la locura y el delirio, de los miedos y sustos que nos asaltan... es ahí donde reside la causa del insomnio y del sonambulismo, de los pensamientos que no llegan, de los deberes olvidados y de las excentricidades...”⁵⁶

Este notable relato, hecho hace 400 años AC, podría haber sido sacado de cualquier libro de neurociencia de este siglo, que señala la implicación del cerebro en la percepción sensorial, el juicio y la emoción, así como su asociación con trastornos mentales. En el libro mencionado, Hipócrates también muestra sus increíbles habilidades anatómicas al proporcionar varias descripciones anatómicas del cerebro humano,

⁵⁶ Publicación original: Chadwick, J. e Mann, N. (eds.) (1983) *Hippocratic Writings*, Penguin: London.

argumentando que es similar al de muchos animales y que está dividido por una estructura que lo separa en dos mitades, una clara referencia al cuerpo calloso y a los dos hemisferios.

Platón también intentó comprender el papel del cerebro en el ser humano, relacionándolo con el concepto de alma. En la filosofía platónica, el término "alma" se distingue del sentido dado normalmente en la religión, ya que posee una estructura tripartita: el "epithymetikon", el "thymos" y el "logistikon". El primero estaba asociado al hígado y los intestinos, donde se encuentran las necesidades vegetativas básicas del individuo, mientras que el "thymos" estaba ubicado en el corazón y provocaba emociones como la ira, el miedo, el orgullo y el coraje.



Imagen 15: linfinitesimale" © Gaia Amoroso

Platón creía que estas dos partes distintas del alma también formaban parte de otros animales y que se extinguían en el momento de la muerte. Sin embargo, el “logistikón” se distinguía de estas dos formas: se consideraba una fuerza espiritual única para los humanos, proporcionando al individuo el pensamiento y la inteligencia, siendo inmortal y capaz de reencarnar.

Ahora, la parte más interesante de esta distinción tripartita es que Platón creía que esta última parte residía en el cerebro. Por ejemplo, en el libro “Fedón”, Platón afirma: “el cerebro es el poder generador de las percepciones del oído, la vista, el olfato, la memoria y la opinión pueden provenir de él”. En el libro “Timeo”, Platón escribe: “la cabeza... es la parte más divina y domina el resto del cuerpo.”⁵⁷

Contrario a esta visión platónica, encontramos nuevamente a Aristóteles. Como mencionamos brevemente antes, el pensamiento aristotélico no concedía mucha importancia al cerebro. Curiosamente, esta conclusión se basaba en datos empíricos y observaciones, aunque fueron interpretados de manera equivocada por el Estagirita.

⁵⁷ Basado en: Crivellato, E. e Ribatti, D. (2007) “Soul, mind, brain: Greek philosophy and the birth of neuroscience”, *Brain Research Bulletin*, 71: 327–336.

Aristóteles estaba fascinado por el funcionamiento del cuerpo humano y fue uno de los primeros seres humanos en disecar una variedad de animales en varios de sus estadios de desarrollo, que incluían peces, reptiles, mamíferos e incluso elefantes.

Aristóteles realizó varias descripciones anatómicas, observando dos membranas que cubrían el cerebro, lo que ahora sabemos que son las meninges ("duramadre" y "piamadre"). También observó una estructura ubicada en la "parte posterior" del cerebro a la que llamó "parencéfalo", que tenía una apariencia muy diferente tanto en textura como en aspecto y que ahora sabemos que se refiere al "cerebelo".

A pesar de estas descripciones precisas, Aristóteles no atribuyó ningún papel relevante al cerebro debido a que no quedó impresionado con su estructura uniforme y fría y relegó mayor importancia al corazón, considerando que este era caliente e irregular. Es importante señalar que una de las creencias fundamentales de esa época estaba relacionada con la tesis de que el calor era esencial para la vida, ya que los cuerpos vivos eran cálidos y los cadáveres fríos.

Curiosamente, esta tesis fundamental tenía bases "empíricas": Aristóteles examinó los diversos estadios embrionarios del huevo de gallina y concluyó que el primer órgano en desarrollarse era el corazón. Entonces, el cerebro era considerado una especie de herramienta biológica para enfriar debido a otra

observación empírica: el hecho de que el calor tiende a subir.

Esto llevó al autor de la *Metafísica* a deducir que los múltiples vasos sanguíneos que cubrían el cerebro servirían para enfriar la sangre del corazón, lo que también le llevó a explicar el tamaño del cerebro humano. Dado que los seres humanos son más cálidos que otros mamíferos y animales en general, necesitaban un "aparato" más grande para reducir la temperatura sanguínea.⁵⁸

Unos pocos años después de estos experimentos rudimentarios de Aristóteles, surgió otro filósofo en la ciudad de Alejandría, llamado Herófilo, quien hizo admirables contribuciones a la anatomía cerebral. Reconoció, por ejemplo, que existirían conexiones muy específicas entre diversas partes del cuerpo y el cerebro a través de la médula espinal, lo que actualmente llamamos "nervios craneales", habiendo descrito siete pares y también formulado sus orígenes: el facial, auditivo, óptico, hipogloso, trigémino y oculomotor.

Así, Herófilo contribuía a la tesis de que era el cerebro - ¡y no el corazón! - el responsable de las funciones mentales en general. Por sus contribuciones verdaderamente notables, fue conocido como el primer gran anatomista, habiendo fundado la "Escuela de

⁵⁸ Basado en: Gross, C.G. (1995) "Aristotle on the Brain", *The Neuroscientist*, 1 (4): 245-250.

Medicina de Alejandría”, donde llevaba a cabo gran parte de las investigaciones con cadáveres.

Después de un paso por el antiguo Egipto y la antigua Grecia, llegamos ahora a la siguiente parada en este viaje histórico, la era del “Coliseo Romano” y el “Imperio Romano”, donde encontramos a Galeno de Pérgamo, considerado el fundador de la fisiología. A diferencia de Herófilo, a Galeno se le prohibió la disección de cadáveres humanos, por lo que utilizó varios animales para llegar a sus descubrimientos.

Galeno logró la hazaña de diferenciar una doble constitución del sistema nervioso. Por un lado, estaba compuesto por vías nerviosas que entraban y salían del cerebro a través de la base del cráneo, por otro lado, también existían un conjunto de vías nerviosas conectadas a través de la médula espinal. Admirablemente, en su obra "Dos Procedimientos Anatómicos", Galeno describe 10 de los 12 pares de nervios craneales que posee cada ser humano, aunque podría haber confundido algunos de ellos entre sí.

Estos descubrimientos, junto con otros avances durante todo el período medieval, allanaron el camino hacia la “Modernidad”, que con Descartes "dividió" el mundo en dos grandes sustancias: la parte material, que comprendía la mayoría de las entidades existentes, y la parte inmaterial, compuesta por el alma o mente humana. Esta forma de concebir el mundo se conoce

como "dualismo de sustancias", ya que divide el universo en una dualidad específica.

Así, Descartes concebiría al ser humano de manera dual en este sentido preciso, el cuerpo y el cerebro estarían constituidos por algo material y físico, mientras que el alma o la mente sería inmaterial y no física. ¿Qué sigue de este esquema dualista? Esta forma de pensar la naturaleza dual del ser humano plantea un problema: el problema de explicar la interacción entre dos sustancias de naturaleza diferente. Si la mente es inmaterial, ¿Cómo podría afectar algo material como el cuerpo o el cerebro?

Descartes enfrentó este problema y siguiendo los pasos de los primeros anatomistas, realizó un estudio muy rudimentario del cerebro humano. Observó que, anatómicamente, la mayoría de las estructuras cerebrales venían en pares (por ejemplo, hemisferio derecho e izquierdo). Sin embargo, una estructura específica parecía ser única: la "glándula pineal". A partir de este descubrimiento, Descartes argumentó que la interacción ocurriría en este órgano particular. Es la glándula pineal la que asegura la interacción causal entre lo mental (imaterial) y lo físico (material).

El autor de la obra "Principios de Filosofía" era, además de filósofo y matemático, tutor de la princesa Isabel de Bohemia, a quien Descartes finalmente dedicaría su libro. Extraordinariamente, Isabel plantearía una de las objeciones más fuertes al dualismo, ¿Cómo puede algo

inmaterial (como el alma) ser afectado causalmente por algo material (el cuerpo y el cerebro)?

¿Cómo podría una creencia (mental), por ejemplo, "quiero dejar de leer este libro", puede llevar al lector a cerrar el libro (un acto físico, corporal)? Esta objeción fue tan contundente que obligó a varios filósofos a presentar otras soluciones para resolver el problema.

Una forma de negar este dilema es aceptar la objeción y decir que, de hecho, no existe una interacción entre lo físico y lo mental. Nicolás Malebranche, filósofo francés, argumentó en este sentido: lo mental y lo físico se relacionan a través de una relación ocasional - a esta posición la llamamos "ocasionalismo". Por otro lado, Gottfried W. Leibniz también argumentó en esta línea, pero de manera diferente a Malebranche: la relación mente-cuerpo ocurriría en una relación de "paralelismo".

El problema con estas dos respuestas a la objeción de Elisabeth es que ambas postulan la existencia de un Dios para sostener el argumento. Por un lado, el ocasionalismo argumenta que Dios es responsable de ordenar lo mental y lo físico constantemente en cada momento: cuando el lector forma la intención de cerrar el libro, Dios interviene y asegura que su cuerpo realice esa acción.

Por otro lado, el paralelismo sostiene que Dios no actúa en cada momento, sino que al principio del universo

creó una sincronización perfecta entre lo mental y lo físico para que parezca que hay una causalidad entre ellos, como si fueran dos relojes sincronizados a la perfección en una armonía preestablecida.

Así, cada vez que el lector forma la creencia "¡esta vez sí que voy a cerrar este libro!", su cuerpo cierra el libro, y esto ocurre no por causalidad, sino por mera sincronización: imagine que la línea de lo mental y la línea de lo físico son dos líneas paralelas que nunca se tocan, pero que están perfectamente sincronizadas entre sí.⁵⁹

Estos problemas con el dualismo han suscitado sospechas sobre él, llevando a filósofos y pensadores a considerar otra posición sobre lo mental y lo físico. La tesis monista de que quizás ambos están constituidos por una misma y única sustancia, la material.

Esta nueva posición abre la posibilidad de estudiar e investigar la naturaleza del cerebro y la mente desde un punto de vista científico, basado en modelos y observaciones rigurosas e informadas, lo que ha llevado a la fundación de una nueva disciplina: la Neurociencia.

Los primeros pasos de esta disciplina comenzaron en 1810 con el proyecto de "Frenología" de Franz-Joseph Gall y J. G. Spurzheim. La Frenología intentaría localizar,

⁵⁹ Basado en: Gouveia, S. (2018) *Philosophical Reflections: Art, Mind and Justice*, Braga: Editora Húmus.

a través de las formas e irregularidades del cráneo humano, diversas funciones, desde el lenguaje hasta la percepción o conciencia.

Sin embargo, este proyecto "localizacionista" fue criticado poco después por el fisiólogo Jean-Pierre Flourens, quien desde Francia rechazó la idea fundamental de que ciertas funciones estuvieran circunscritas a regiones específicas del cerebro. Por otro lado, Hughlings Jackson, neurólogo en Inglaterra, argumentó que, basado en su trabajo con pacientes epilépticos, diversas funciones estarían asociadas con diversas regiones del cerebro humano.

Un poco más tarde, en 1861, surgió el famoso caso reportado por Paul-Pierre Broca que parecía corroborar la tesis de que una lesión en una área específica del cerebro -en el lóbulo frontal izquierdo, que posteriormente se conocería como el "Área de Broca"- estaba relacionada con un trastorno particular llamado afasia, que afecta la habilidad para comprender y expresar el lenguaje.

En Alemania, Karl Wernicke se especializó en este tema, obteniendo su doctorado en 1876 con una tesis centrada en una víctima de derrame cerebral que había afectado otra región relevante del cerebro para el lenguaje, que posteriormente sería conocida como el "Área de Wernicke".

Un poco más tarde, en Italia, surge Camillo Golgi, quien hizo una importante contribución al campo de la neurociencia, desarrolló una técnica específica para identificar la estructura de las neuronas que, lo que permitió por primera vez la visualización de neuronas.

El método de Golgi implicaba impregnar el tejido cerebral con cromato de plata, lo que permitía la visualización de neuronas individuales al teñir un número limitado de células en su totalidad. Este avance proporcionó una nueva perspectiva sobre la estructura intrincada de las neuronas, incluyendo sus dendritas, axones y cuerpos celulares.

Utilizando el mismo método, Santiago Ramón y Cajal, en España, contribuyó con la idea de que las neuronas eran estructuras unitarias que transmitían información solo de las dendritas a los axones, lo contrario de lo que se pensaba hasta entonces. Estos descubrimientos, junto con otros, llevaron a filósofos y científicos a formular una conexión cada vez más precisa entre aspectos mentales y partes específicas del cerebro.

El estudio de las enfermedades neurológicas también jugó un papel crucial en el avance de la neurociencia. Erwin Strauss, Kurt Goldstein y Hartmut Kuhlenbeck se dedicaron al análisis de los cerebros "enfermos" de varios soldados de la "Primera Guerra Mundial". Inicialmente, se centraron en la localización específica de las lesiones y examinaron el impacto en el

comportamiento y las funciones asociadas a esas lesiones en dichos soldados.

Curiosamente, en un enfoque único, interrogaron a los pacientes sobre cómo sus lesiones estaban influenciando sus experiencias corporales y sus relaciones con el mundo, estableciendo así una conexión indirecta (por primera vez) entre las descripciones fenomenológicas de la conciencia y las características neuronales del cerebro.

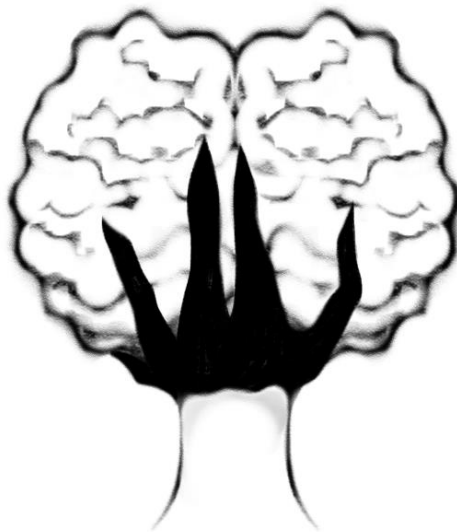


Imagen 16: Físico e mental” © Gaia Amoroso

El avance en el desarrollo de tecnologías específicas para explorar directamente el cerebro también fue

fundamental para el desarrollo de varios conocimientos clave sobre el sistema nervioso. Las investigaciones del alemán Hans Berger permitieron registrar por primera vez la actividad eléctrica en el cráneo de seres humanos.

Esta técnica fue mejorada a lo largo de varios años hasta que, en la década de 1930, Berger desarrolló la técnica de electroencefalografía (mejor conocida como EEG), que permitía la identificación de varios ritmos cerebrales específicos. Por ejemplo, a frecuencias más bajas encontramos la onda delta, que está en el rango de 0.5-4 Hz y está asociada con etapas de sueño profundo, mientras que a frecuencias más altas encontramos la onda gamma, que está en el rango de 30-100 Hz y está asociada con procesos superiores como la memoria, la percepción o el aprendizaje.

Por último, en las últimas décadas se han desarrollado otras tecnologías relevantes para el estudio del cerebro, que incluyen la tomografía por emisión de positrones (PET) y la resonancia magnética funcional (fMRI). Estas tecnologías no miden directamente la actividad eléctrica, sino que evalúan la actividad bioquímica o metabólica en las neuronas.

Por ejemplo, utilizando fMRI, podemos cuantificar con cierta precisión la cantidad de moléculas de oxígeno consumidas por las neuronas y, basándonos en ese dato, inferir el nivel de la actividad eléctrica. Por otro lado, utilizando PET, utilizamos sustancias para trazar

receptores específicos o diferentes balances bioquímicos en las neuronas.

A distinción fundamental entre las tecnologías de EEG y fMRI/PET radica en los tipos de resoluciones que pueden alcanzarse. En el caso del EEG, se logran mediciones de tiempo de alta resolución en el orden de milisegundos, pero con baja resolución espacial. Por otro lado, con fMRI/PET obtenemos una alta resolución espacial, pero una baja resolución temporal en el orden de segundos. Debido a esta distinción, muchos estudios actuales en neurociencia intentan crear estudios combinados utilizando tanto EEG como fMRI/PET para obtener una interpretación más rigurosa de los datos.⁶⁰

Después de esta breve odisea histórica sobre el cerebro y las reflexiones filosóficas sobre su naturaleza, composición y relevancia, en esta segunda parte procederemos a introducir otras aproximaciones interesantes sobre la naturaleza del cerebro con cuatro pensadores de renombre internacional (Anil Seth, Karl Friston, Christof Koch y Joseph LeDoux) con quienes he tenido el privilegio de dialogar.

Comencemos entonces esta segunda parte, dedicada al cerebro, introduciendo una visión más contemporánea de este importante órgano que sostiene que, contrariamente a lo que pensábamos hasta ahora, el

⁶⁰ Basado en: Gouveia, S. (2022) *Philosophy and Neuroscience: a Methodological Analysis*, New York: Palgrave Macmillan.

cerebro puede ser visto como una "máquina de predicción" cuya función es crear hipótesis sobre el mundo.

II. EL CEREBRO PREDICTIVO

El primer diálogo de esta segunda parte, centrado en el cerebro, cuenta con la participación del neurocientífico que popularizó la idea de considerar el cerebro como una "máquina de predicción". Tiene una de las charlas TEDx más vistas del mundo, con 10 millones de visualizaciones y lleva el sugestivo título "Tu cerebro alucina tu realidad consciente".⁶¹

Hablamos de Anil Seth, profesor de "Neurociencia Cognitiva y Computacional" en la Universidad de Sussex en Inglaterra, donde también es codirector del "Sackler Center for Consciousness Science". Seth defiende una teoría conocida como "Procesamiento Predictivo" (PP).

Esta aproximación surge como una crítica o respuesta a una teoría que atribuye al cerebro un papel únicamente pasivo, es decir, el de recibir estímulos del exterior (¡de la realidad!) y, a partir de esos estímulos, construir un mapa o modelo mental de esa realidad, sin considerar relevante un papel activo del cerebro.

Por lo tanto, esta nueva teoría critica esa postura pasiva y sostiene que el cerebro es un mecanismo de predicción perfectamente activo cuya función es

⁶¹ YouTube: <https://www.youtube.com/watch?v=lyu7v7nWzfo>.

anticipar los estímulos del mundo para combinarlos entre sí y formar la mejor "suposición" de lo que podría haber causado esos mismos estímulos.

Pensemos en la percepción visual. Contrariamente a lo que se defendía hasta entonces, la percepción se veía como un mecanismo donde el lector recibía estímulos visuales de las palabras que está leyendo en este libro. Solo después de recibir esas señales es que su cerebro construía una representación mental del libro.

El "Procesamiento Preditivo" sostiene que esta interpretación es incorrecta: la percepción tiene una naturaleza constructiva que está en una danza constante con las señales recibidas del mundo. Según esta teoría, la percepción es una forma de "alucinación controlada" que intenta prever un modelo de las señales visuales que el lector va a recibir y, solo después, verifica en la realidad si esas señales son correctas o si necesitan ser actualizadas por otras señales.

Solo después de esta predicción es cuando el cerebro confirma si esas señales son precisas o si necesitan ser actualizadas por señales adicionales. Si esta explicación parece un poco confusa, animo al lector a dirigir su atención a la siguiente imagen:



Imagen 16: Ilustración del Procesamiento Predictivo (percepción visual) I

Si el lector nunca ha visto esta imagen, podría encontrar que tiene poco sentido, pareciendo en el mejor de los casos una pieza de gusto estético dudoso que podría haber sido producida por un bebé al estilo de Pollock. No hay nada en el mundo que podamos identificar con esas formas y sombras.

Sin embargo, aunque su cerebro no pueda identificar nada, el enfoque defendido por Anil Seth se vuelve más plausible si ahora miramos la segunda imagen, que tiene apenas un pequeño detalle diferente:

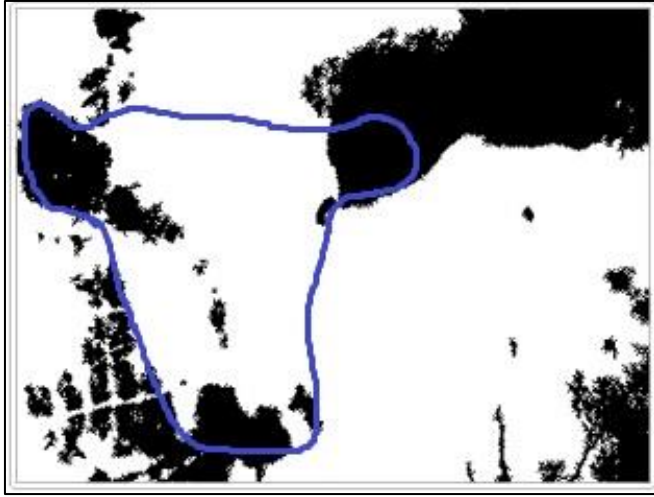


Imagen 17: Ilustración del Procesamiento Predictivo
(percepción visual) II

En esta segunda imagen, al añadir una línea que indica parte de las formas y sombras que antes carecían de significado, su cerebro ahora puede comprender lo que está presente. La segunda imagen permite que el cerebro actualice el modelo anterior y cree uno nuevo en el que puede detectar la forma rudimentaria de la cara de una vaca. Lo interesante de esta "experiencia" visual es que, si el lector vuelve a ver la primera imagen, ahora será capaz de percibir la cara de la vaca sin la ayuda de la línea, algo que antes su cerebro no pudo ver.

Pero ¿Cómo explica el “Procesamiento Predictivo” este fenómeno? En esencia, la idea es que, después de que su cerebro percibió la segunda imagen, procedió a actualizar el modelo anterior que antes carecía de significado, incorporando esta nueva información al modelo previo. Esto permite que algo que antes parecía no tener ningún significado ahora lo tenga.

Note que la información sensorial que su cerebro tiene tanto la primera vez que mira la primera imagen, como cuando la mira por segunda vez (después de ver la segunda imagen), es exactamente la misma: el estímulo visual no ha cambiado, se ha mantenido igual. Lo que ha cambiado es el mejor "supuesto" que su cerebro considera plausible basado en experiencias previas.

En este sentido, esta teoría sostiene que lo que vemos ocurre tanto por la influencia del mundo en el cerebro, pero también, y posiblemente más importante aún, por la influencia del cerebro en la predicción de ese mundo.

Es importante notar que el “Procesamiento Predictivo” puede aplicarse a la percepción y los sentidos en general, pero también a muchos otros aspectos de la mente. En esencia, este enfoque teórico intenta explicar la naturaleza de todos los aspectos mentales, desde la conciencia hasta la memoria e incluso las enfermedades psiquiátricas, basándose en el mismo principio fundamental. A continuación, veamos otro ejemplo interesante para mostrar el potencial de este marco aplicado al “yo” (al sí mismo).

La ilusión de la mano de goma es un fenómeno estudiado en Psicología que se centra en cómo su cerebro predice, al igual que la percepción, su “yo”, es decir, lo que su cerebro considera parte de su cuerpo. Esta experiencia⁶² tiene la siguiente estructura:

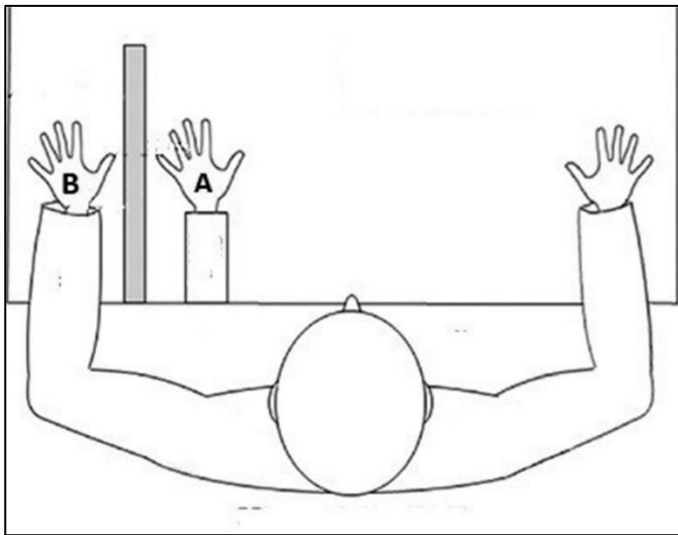


Imagen 18: Ilustración del Procesamiento Predictivo en el “yo”; Mano A = de goma/falsa; Mano B = verdadera.

Imagínese que usted es la persona sentada con una mesa frente a usted, lista para participar en esta experiencia. Lo que va a suceder es lo siguiente: su

⁶² Un ejemplo de la experiencia en vídeo se puede encontrar en el siguiente enlace:
https://www.youtube.com/watch?v=nzF_DfOafKw

mano izquierda estará apoyada sobre la mesa, pero usted no podrá verla directamente porque algo está bloqueando su acceso visual a ella (en la imagen, estará en la posición B). Un poco más a la derecha de su mano izquierda estará una mano de goma que usted puede ver y que está colocada donde debería estar su mano izquierda real si no estuviera bloqueada (en la imagen, en la posición A).

El próximo paso del experimento es tener a una persona frente a usted que tocará simultáneamente su mano real (que no puede ver porque está tapada) y la mano de goma (que puede ver frente a usted). Finalmente, después de unos minutos de estimulación, algo mágico parece suceder. Michael Myers aparece de repente y apuñala la mano de goma que está frente a usted, automáticamente y de inmediato retirará su mano izquierda real de la mesa, actuando como si hubiera sido su propia mano la apuñalada en lugar de la mano de goma.

Lo que sucede aquí es que la información sensorial del tacto, combinada con la información visual, es suficiente para que su cerebro cree un modelo de su cuerpo en el cual la mano de goma se percibe como parte de él. Desde la perspectiva de esta teoría, incluso la experiencia de lo que pertenece a nuestro cuerpo es una "alucinación controlada" por el cerebro, es decir, una suposición o predicción.

Ahora bien, al igual que podemos ser engañados por ilusiones visuales, la forma en que experimentamos nuestro cuerpo también puede llevar a modelos erróneos. Esta teoría podría ofrecer una contribución importante para entender enfermedades psiquiátricas que involucran trastornos del “yo” como el trastorno bipolar o la esquizofrenia.

¿Pero de qué manera se lleva a cabo esta continua predicción que parece estar haciendo nuestro cerebro según esta teoría, tanto en relación con los señales externas de la realidad como con los señales internos de nuestro cuerpo?

Un concepto fundamental para entender este enfoque es el de “Inferencia Bayesiana”, que permite explicar cómo, por un lado, formamos nuestros modelos predictivos y, por otro lado, cómo podemos actualizarlos constantemente y ajustar nuestras creencias a nueva información.⁶³

En la práctica, la inferencia bayesiana⁶⁴ implica combinar dos fuentes de conocimiento: las predicciones generadas internamente por el cerebro, a

⁶³ Basado en: Gouveia, S. & Curado, M. (2020) (eds.) *The Philosophy and Science of Predictive Processing*, New York: Bloomsbury.

⁶⁴ En su forma más formal, la probabilidad posterior $[P(H | E)]$ se calcula mediante la conjunción de la verosimilitud $[P(E | H)]$ con la probabilidad anterior $[P(H)]$, dividida por la verosimilitud marginal $[P(E)]$, donde E representa la Evidencia, es decir, los nuevos datos del mundo, y H representa una Hipótesis cuya probabilidad puede ser influenciada por E.

las que llamamos "anterior", y la información sensorial proveniente de la realidad, a la que llamamos "verosimilitud". Estas dos fuentes se combinan para formar una nueva estimación, a la que llamamos "posterior", que representa el modelo actualizado del mundo. En esencia, nuestro cerebro busca equilibrar sus expectativas internas con la evidencia externa para adaptarse constantemente a la realidad.

Veamos un ejemplo simple que demuestra la utilidad de este método. Empecemos con el "anterior". Antes de entrar en casa, el lector asume que hay una probabilidad mínima (1%) de encontrar un gato, dado que sabe que no tiene ningún gato. Sin embargo, al abrir la puerta, escucha un maullido (una observación, que se ajusta a la "verosimilitud") y, sabiendo que los gatos maúllan, el lector actualiza su creencia de que podría haber una alta probabilidad (90%) de que haya un gato en su casa. Al combinar esta creencia inicial con la nueva evidencia, se llega al "posterior". Hay una alta probabilidad de que hay un gato en su casa.

Otros conceptos fundamentales para entender los fundamentos de la teoría del "Procesamiento Predictivo" como una teoría unificadora de la mente humana, como la "Inferencia Activa", el "Error de Predicción" y el "Principio de la Energía Libre", se introducirán en la próxima sección. Antes de eso, veamos también la utilidad de este enfoque para

explicar la naturaleza de diversas enfermedades psiquiátricas.

Siguiendo lo dicho anteriormente, el cerebro es una máquina de predicciones que combina modelos internos con estímulos del mundo. Podemos argumentar que, en la esquizofrenia, esta capacidad de realizar predicciones de manera más efectiva se ve comprometida por varias razones. ¿Cuáles podrían ser esas razones?

Sabemos que uno de los síntomas positivos más comunes en esta enfermedad son las alucinaciones auditivas, donde la persona cree que las "voces" generadas internamente tienen un origen externo en el mundo, es decir, que alguien realmente les está diciendo algo.



Imagen 19: "Rifiuto della coscienza" © Gaia Amoroso

Los individuos con autismo pueden tener dificultades para integrar información sensorial externa, lo que puede llevarlos a concentrarse demasiado en sus propios modelos internos.

Dado que estos individuos tienen más dificultad para incorporar un nuevo estímulo sensorial, su capacidad para actualizar los modelos y, por lo tanto, para adaptarse de manera más efectiva al mundo, queda seriamente comprometida.

Es sabido que estos individuos prefieren mantener rutinas y patrones particulares, y tienen grandes dificultades en las relaciones sociales precisamente porque tienen una capacidad reducida para prever el comportamiento de los demás.

Imagine ahora a una persona cuyo cerebro tiene una tendencia sesgada de manera negativa para anticipar los modelos del mundo. Esto llevaría a esa persona a tener expectativas constantemente pesimistas, distorsionando su capacidad para prever resultados positivos.

Esto puede basarse en una "rigidez" en la relación del "yo" con el mundo, haciendo que el modelo relacional entre ellos permanezca negativamente rígido, es decir, no permitiendo la entrada de nueva información positiva y perpetuando el sesgo negativo.

En este caso, estaríamos ante un paciente con... depresión, una enfermedad que afecta a millones de personas en todo el mundo. Sabemos, por ejemplo, que las personas con depresión tienen una tendencia a verse a sí mismas de manera negativa, procesando solo la información que confirma una baja autoestima sobre ellas mismas.

Y si un cerebro otorga una importancia predictiva excesiva a estímulos, generando constantemente predicciones de eventos futuros considerados amenazantes, lo que lleva a una anticipación exagerada que resulta en preocupaciones excesivas por el peligro. ¿Qué problema psiquiátrico tendría una persona así? Exactamente: ¡Ansiedad (en este caso, crónica)!

Dado que existe una interpretación distorsionada de diversas amenazas en el ambiente, el cerebro de un individuo ansioso tendrá dificultades, por un lado, para actualizar sus modelos internos de "seguridad" en su entorno y, por otro lado, amplificará estímulos cotidianos como posibles peligros, lo cual contribuirá a su sintomatología de ansiedad.

Como podemos percibir a través de estos ejemplos de Psiquiatría, la teoría del "Procesamiento Predictivo" tiene un enorme potencial explicativo y podría efectivamente contribuir al descubrimiento de mecanismos específicos en diversas enfermedades mentales. Además, también podría explicar cómo

funciona nuestra percepción, nuestro “yo”, nuestra memoria, los sueños e incluso la conciencia.

A continuación, y en continuación con este enfoque, introduciremos algunas ideas que sirven de base para esta teoría, desarrolladas por uno de los neurocientíficos más influyentes del mundo, el Profesor Karl Friston.

III. EL CEREBRO LIBRE

El segundo diálogo cuenta con la participación del profesor Karl Friston, una de las mayores autoridades en las ciencias del cerebro y uno de los fundadores de la perspectiva descrita en la sección anterior. Después de una breve introducción al "Procesamiento Predictivo", presentaré a continuación algunos principios que forman la base de este enfoque del cerebro y la mente humana, y esta sección estará en diálogo directo con la anterior.

Comencemos por introducir el "Principio de la Energía Libre". Este principio tiene una larga historia y ha sido desarrollado por varias generaciones de pensadores de diversas disciplinas, habiendo sido formalizado más recientemente en la neurociencia por Karl Friston, el científico vivo más citado del mundo. En esencia, se sugiere que el cerebro humano tiene una propensión intrínseca a minimizar el consumo de energía al realizar diversas tareas cognitivas y mentales en general.

La idea es que el cerebro, para economizar energía, busca organizarse de la forma más eficiente posible, optimizando las predicciones sobre el mundo que describimos anteriormente. Esto significa que, según este principio, el cerebro busca crear modelos internos precisos del mundo con el mínimo "esfuerzo" posible.

En términos prácticos, el “Principio de la Energía Libre” implica pensar en el cerebro como si estuviera constantemente buscando un equilibrio entre, por un lado, mantener los modelos internos estabilizados y, por otro, ser lo suficientemente flexible para adaptarse a la nueva información del mundo, permitiendo que esto ocurra de la manera más económica y eficaz posible. Es un enfoque que permite formular cómo los organismos manejan la incertidumbre.

Debemos notar que este principio tiene una forma matemática compleja que no podremos detallar en un libro introductorio de esta naturaleza. Pero, en esencia, se trata de un principio que incluso puede ser considerado "simple" de comprender. Imagine que el lector está conduciendo por una carretera que le es familiar (por ejemplo, del trabajo a casa, algo que hace todos los días).



Imagen 20: “Libertà” © Gaia Amoroso

En ese caso, su cerebro ha construido, a lo largo de los numerosos viajes realizados, un modelo interno eficaz del recorrido que debe seguir. Esto incluye la ubicación de las curvas, los semáforos, las señales de tráfico y otros puntos de referencia, en un intento de economizar la energía necesaria para prever cómo será la carretera la próxima vez que la utilice.

Ese modelo interno utiliza, por lo tanto, la información recopilada en el pasado y usa esa misma información para prever lo que podría suceder, por ejemplo, si el lector ya sabe que hay un semáforo al final de la calle, su cerebro anticipará ese hecho y no se sorprenderá si el semáforo se pone en rojo. Esto permite la creación de un modelo "estable" de la carretera, que se actualiza solo si es necesario, lo que permite un considerable ahorro de energía.

Sin embargo, si por algún motivo ocurre alguna alteración (por ejemplo, se instala un nuevo semáforo), su cerebro buscará adaptarse e incorporar esa nueva información al modelo ya existente. En lugar de reformular todo el modelo interno de la carretera, actualizará el anterior con ese nuevo dato, procurando constantemente disminuir los elementos de sorpresa.

A pesar de este ejemplo simple, es necesario comprender la fuerza de este principio. En esencia, es lo que nos permite explicar cómo un organismo se

mantiene vivo y se aplica tanto a células simples como bacterias, como a organismos altamente complejos como el ser humano y otros animales.

Tanto los organismos simples como los desarrollados son considerados sistemas dinámicos por naturaleza, separados entre su propia existencia y un entorno donde esa existencia ocurre, lo que implica una interacción indirecta y constante intercambio de información. El propósito de esta interacción es llegar a un segundo concepto fundamental, el concepto de "Minimización del Error de Predicción".

El objetivo de lo discutido hasta ahora es garantizar que los modelos predictivos creados por el cerebro minimicen la posibilidad de error. Un modelo preciso estará más cerca de la realidad y, por lo tanto, no requerirá un gasto energético tan costoso en comparación con un modelo menos preciso que necesitará ser actualizado.

Así, todos los organismos intentan minimizar los errores de las predicciones sobre el entorno a través del "Principio de la Energía Libre". ¿Qué significa "energía libre" en este contexto? La energía libre puede considerarse una métrica relacionada con la sorpresa o la incertidumbre en sus predicciones. Cuanto más un organismo logre minimizar su energía libre, más podrá garantizar su supervivencia.

Es importante notar que este principio actúa tanto en la predicción de los estados externos del ambiente (como ya hemos mencionado), como también en la predicción de los estados internos del cuerpo. El organismo, además de prever modelos de cómo será el ambiente en un momento dado, al mismo tiempo está prediciendo las condiciones que los órganos internos deben producir para garantizar la supervivencia en ese ambiente.

Esto podría implicar lo siguiente: si hay errores en la predicción de ciertas condiciones externas (del ambiente) o internas (del cuerpo), nos encontraremos en un escenario que conduce a una alta energía libre. Esto, a su vez, se manifestará en estados de sorpresa o alta imprevisibilidad.

Si, en cambio, la predicción es precisa, el resultado será una baja energía libre. Por lo tanto, todos los organismos, sean complejos o simples, mamíferos o aves, plantas o bacterias, buscan minimizar esa incertidumbre o sorpresa.

Es importante señalar otra característica de este principio. Muchas teorías biológicas -aquellas que intentan explicar por qué un fenómeno biológico funciona de cierta manera- intentan responder a una pregunta que este enfoque intenta invertir, en lugar de enfocarse en describir qué debe hacer un organismo para existir, se busca comprender, asumiendo que el

organismo ya existe, qué necesita hacer para continuar existiendo.

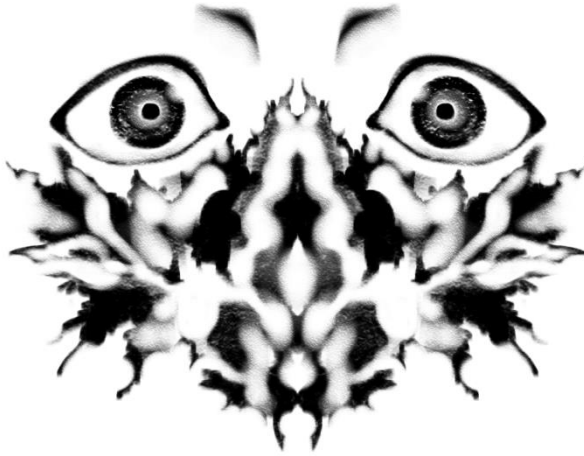


Imagen 21: "lo esisto" © Gaia Amoroso

Esto es relevante para entender que el "Principio de la Energía Libre" en sí mismo no es una teoría falsable sobre cómo se comportan los organismos, solo sus postulados pueden ser falsables o no, dependiendo de su aplicabilidad.

Finalmente, para concluir esta breve introducción, nos queda abordar otro concepto fundamental, el de la "Inferencia Activa". Este concepto permite que el organismo, cuando el cerebro detecta algún problema en los modelos generados en cierto contexto, busque

nueva información relevante para esa situación particular con el fin de ajustar el modelo predictivo incorrecto o impreciso.

Así, la “Inferencia Activa” puede ser considerada como la capacidad del cerebro de influir activamente en la información que recogemos del entorno, lo que implica seleccionar ciertos estímulos y dirigir la atención hacia aquellos factores (ya sean externos o internos, nuevamente) que pueden ser relevantes para reducir la ambigüedad o incertidumbre identificada.

Veamos un ejemplo relacionado con la predicción de estados internos del cuerpo, como los estados fisiológicos. Imagina que de repente te trasladan desde la playa de Copacabana a los increíbles Alpes Suizos. Tu cuerpo, a través de la interocepción, detecta un cambio de temperatura y comienza a producir una sensación de frío. Este señal interoceptiva indica que la temperatura de tu cuerpo disminuirá si te mantienes en ese ambiente.

Lo que permite la “Inferencia Activa” en este escenario es buscar información sobre ese ambiente y también sobre otros señales internos del cuerpo que podrían haber sido activados. Por ejemplo, tu cerebro podría haberse enfocado en los temblores que tus músculos comenzarían a experimentar, lo cual sería un indicador de la temperatura real de tu cuerpo.

Basándose en todos estos elementos, tu cerebro reformula el modelo de temperatura que tenía en Copacabana (donde, por ejemplo, estarías sudando para combatir el calor), hacia un modelo en el que otros elementos, como los temblores (que tienen el propósito de aumentar el calor interno), pasan a formar parte del mismo.

Por lo tanto, es a través de la “Inferencia Activa” que, en este ejemplo, el cerebro optimizó activamente – enfocando su energía y atención tanto en las señales externas del ambiente como en las señales internas del cuerpo– las predicciones anteriores para mantener una regulación adecuada de la temperatura corporal que, si no hubiera sucedido, podría haber llevado a consecuencias... un poco desagradables.

Algunos neurocientíficos y filósofos han criticado este enfoque,⁶⁵ argumentando que no todo lo que sucede en el organismo parece tener como objetivo minimizar el consumo de energía, dado que hay varias funciones mentales, como la creatividad o la imaginación, que parecen exigir un gasto considerablemente mayor de lo normal.

La propia variación neuronal interindividual (entre personas) parece diferir y esta diversidad parece ser

⁶⁵ Article by Professor Karl Friston: Friston, K. et al. (2023) “The free energy principle made simpler but not too simple”, *Physics Reports*, 1024: 1-29.

difícil de explicar apelando a la actuación del mismo principio. Pero el mayor desafío quizás sea medir y cuantificar de hecho todo el mecanismo predictivo que ocurre tanto en el cerebro como en cada célula.

A continuación, abordaremos otra perspectiva para observar el cerebro y su papel en el desarrollo de procesos conscientes, presentando una teoría que pretende proporcionar precisamente una forma de cuantificar la mente humana. Esta teoría fue desarrollada por el neurocientífico y presidente del “Allen Institute for the Brain en Seattle”, el Profesor Christof Koch.

IV. EL CEREBRO INTEGRADO

En este tercer diálogo, introduciremos una de las teorías neurocientíficas más en boga sobre el cerebro y su papel en el desarrollo de la mente consciente, desarrollada por el Profesor Christof Koch, autor del libro "The Feeling of Life Itself: Why Consciousness Is Widespread but Can't Be Computed".

Esta teoría, conocida como "Teoría de la Información Integrada", (TII),⁶⁶ ha revolucionado la forma en que la conciencia es concebida dentro de la neurociencia. Por su carácter conceptual y teórico, ha impactado varias áreas adyacentes, como la Filosofía, la Física y la Inteligencia Artificial. Debido a su relevancia, esta teoría también ha dado lugar a consideraciones éticas interesantes.

En primer lugar, es importante señalar que este enfoque busca proporcionar una respuesta científica y objetiva al problema causado por la existencia de la conciencia subjetiva en el mundo físico. La teoría

⁶⁶ El primer autor en desarrollar esta teoría fue Giulio Tononi en 2004, con Christof Koch, después de haber trabajado con el ganador del Premio Nobel Francis Crick, se unieron para seguir desarrollando esta teoría. Publicación original: Tononi, G. (2004) "An Information Integration Theory of Consciousness", *BMC Neuroscience*, 5 (1): 42.

intenta identificar las propiedades esenciales, llamadas "axiomas", e inferir las condiciones necesarias y suficientes que cualquier sustrato debe cumplir para ser consciente, denominadas "postulados". Todo esto se expresa a través de una notación matemática específica que, debido a su complejidad, no podremos abordar en este libro.⁶⁷

La estrategia de la TII busca cambiar el enfoque típico que se encuentra en la neurociencia en la investigación de la conciencia. En lugar de intentar identificar en el cerebro los procesos neuronales responsables de los estados conscientes, Koch prefiere comenzar desde la fenomenología (¡desde la conciencia!) y luego preguntar qué tipo de mecanismos físicos podrían explicar esa fenomenología de la experiencia.

Para este enfoque, la definición de experiencia consciente está muy cercana a la ofrecida por Descartes, quien consideraba la conciencia como una propiedad fundamental del mundo. Partiendo de este supuesto, la TII avanza con los siguientes 5 axiomas:

- Axioma de la Existencia (AExis): este axioma es, en esencia, una reformulación del cogito

⁶⁷ Si estás interesado, puedes encontrar una descripción matemática de la TII en el siguiente artículo (disponible gratuitamente en línea): Albantakis L. et al. (2023) "Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms", PLOS Computational Biology, 19 (10): e1011465.

cartesiano, sustituyendo el acto de "pensar" con el acto de "existir": "experimento, luego existo". Esto significa que tenemos una certeza indudable e inmediata de nuestra propia perspectiva en primera persona en el mundo; según este axioma, la conciencia es intrínsecamente real;

- Axioma de la Información (AInfo): este axioma indica que una experiencia siempre debe especificar algo, siendo siempre distinta de otras experiencias posibles; por lo tanto, podemos afirmar que cada experiencia nos informa de manera particular a través de un contraste con otras experiencias particulares;
- Axioma de la Integración (AInteg): este axioma sostiene que la conciencia es unificada y que el contenido de la experiencia no se reduce a partes independientes, sino que está integrado como un todo; por lo tanto, no tenemos experiencias aisladas unas de otras que luego se forman como un conjunto mental; más bien, cuando el lector experimenta la lectura de estas palabras, experimenta el tipo de letra y el color en una experiencia unificada, no separados entre sí;
- Axioma de la Composición (AComp): este axioma sostiene que todas las experiencias tienen una estructura compuesta por varios aspectos y combinaciones entre ellos; nuevamente, si el

lector está experimentando estas palabras, estas contienen diferentes aspectos fenomenológicos como los colores y las formas de las letras.

- Axioma de la Exclusión (AExcl): este último axioma indica que la experiencia consciente siempre excluye otras experiencias, ya que al especificar una experiencia (cf. AInfo), necesariamente debe excluir otras experiencias; en otras palabras, las experiencias conscientes tienen fronteras definidas; además, tienen un "grano" temporal: es decir, los contenidos de la experiencia tienen una duración específica en el tiempo.

Vamos ver cómo estos axiomas pueden aplicarse a un ejemplo cotidiano, como el de leer un libro, siguiendo el orden de los axiomas presentados.



Imagen 22: "No brain, never mind" © Gaia Amoroso

Imagina que estás en casa, en el sofá, disfrutando de tu preciado tiempo libre, hojeando entusiastamente este libro, lo cual indica la presencia de experiencias conscientes intrínsecas (AExis). Durante esa lectura, la información del libro se procesa de manera única en tu conciencia. Los diversos patrones y conocimientos contenidos en las palabras demuestran el carácter informativo de esa experiencia (AInfo).

Además, esta información presente en el libro forma una experiencia consciente unificada e integrada de forma coherente entre las palabras, las imágenes y las emociones que se pueden sentir al leer el libro (AInteg). Por supuesto, esta experiencia de lectura consciente se compone de elementos particulares, como las distintas discusiones y las introducciones del autor: todos estos elementos contribuyen a una composición única de su experiencia de lectura.

Finalmente, mientras el lector está leyendo este libro, pueden estar ocurriendo muchas otras cosas simultáneamente, como un partido de fútbol de su equipo favorito en la televisión. Sin embargo, si considera interesante el contenido de este libro, su experiencia consciente podría incluso excluir todas las demás experiencias simultáneas, demostrando la naturaleza única de la conciencia en un momento dado (AExcl).

Además de estos axiomas, que se consideran verdades autoevidentes, la TII postula, basándose en cada uno de

estos axiomas, qué tipo de propiedades debe poseer un sistema físico para ser consciente. Ahora veamos los postulados de la TII:

- Postulado de la Existencia: Este postulado afirma que la existencia de la consciencia implica un sistema de mecanismos con poder de causa y efecto; para que exista un sustrato físico, una condición necesaria es tener poder causal; por ejemplo, la neurona es un mecanismo de estados binarios, ya que puede describirse como "disparando" o "apagada";
- Postulado de la Información: Este postulado declara que, si la conciencia es realmente informativa, debe tener la capacidad de especificar o diferenciar ciertas experiencias de otras; esto implica que cualquier mecanismo dentro de un sistema físico debe tener poderes de causa y efecto; todos estos "reportorios" forman la estructura causa-efecto que permite al sistema especificar un estado determinado;
- Postulado de la Integración: Este postulado indica que, para que la integración de la conciencia ocurra en un todo unificado, el sistema físico debe ser necesariamente irreducible: las partes de este sistema deben ser interdependientes; todos los elementos mecánicos del sistema deben ser capaces de causar algo en el todo, así como ser afectados por él; si un sistema físico puede ser dividido en

partes sin que esto afecte su estructura causa-efecto, entonces dicho sistema pierde su integración y no puede ser consciente.

- Postulado de la Composición: Este postulado implica asumir que los elementos (mecanismos) de un sistema físico deben tener la capacidad de combinarse entre sí, garantizando que esas combinaciones posean poder causa-efecto; es decir, si una experiencia tiene una estructura determinada, el proceso causal que genera esa experiencia debe ser necesariamente estructurado.
- Postulado de la Exclusión: Finalmente, este postulado implica la idea de que un estado consciente de un sistema físico debe ser finito: diversos mecanismos de un sistema pueden tener diversos poderes causa-efecto; sin embargo, para que haya conciencia en un sistema, solo uno de estos mecanismos puede tener la estructura causa-efecto máximamente irreductible, lo que representa el nivel más elevado de integración en términos de información.

Veamos otro ejemplo para entender la relevancia de lo que se acaba de describir. Imagina que el lector se despierta repentinamente de su sueño profundo y ve frente a él la pared azul de su habitación. Además, también ve dónde está su cama y se da cuenta de que,

en ese momento, está consciente de sí mismo y de lo que lo rodea.

Ahora bien, lo que la TII afirma es que esta experiencia existe por sí misma y no por cualquier otra cosa; es una experiencia específica (no genérica); es una experiencia unitaria. Tu ojo derecho no está experimentando algo diferente de tu ojo izquierdo (ambos experimentan la misma experiencia); además, esta experiencia está definida específicamente por lo que estás observando frente a ti y no por lo que está detrás de ti o a tu lado. Finalmente, esta experiencia está estructurada por varios elementos que se relacionan entre sí (por ejemplo, tu cuerpo, en tu cama, situada de tal manera que puedas ver la pared azul frente a ti), lo que te permite tener precisamente esta experiencia y no otra.

¿Has notado lo que acabamos de hacer? Exactamente eso. Mediante este ejemplo, aplicamos todos los postulados de la TII.

Es importante destacar que, para los teóricos de la TII, estos axiomas (y por lo tanto, los postulados) forman parte de una lista completa y finita. No hay ninguna otra propiedad de la experiencia consciente que se considere esencial para su existencia. Si un sistema físico —sea biológico, artificial o incluso alienígena— desea ser consciente, deberá reproducir a través de los postulados todos los axiomas presentados aquí.

Otro aspecto atractivo de esta teoría es proporcionar un intento de cuantificar, en una medida particular, la capacidad de un sistema para integrar información. Esta cuantificación se presenta a través de la medida Φ —la letra griega Phi— que representa la habilidad (entre cero y máxima) de un sistema para integrar la información procesada en un sistema particular.

Según la TII, la conciencia "ocurre" cuando un sistema alcanza un alto nivel de información integrada. Cuanto más intrínseca e irreductible sea esta información, mayor será su Φ y, por lo tanto, mayor será el grado de conciencia. El cálculo se realiza mediante la cantidad de información que está globalmente integrada en ese sistema, comparada con la suma de la información de cada una de sus partes.

Entonces, siempre que un sistema alcanza un Φ máximo, se considerará que ese sistema es consciente. Por debajo de ese valor máximo, se sostiene que el sistema no tiene suficiente información integrada para alcanzar un estado consciente.

Este enfoque cuantificable, junto con los axiomas y los postulados de la TII, proporciona una teoría increíblemente elegante y, además, con la capacidad de poner a prueba sus supuestos. Veamos algunas predicciones interesantes, basadas en datos empíricos reales, sobre el poder explicativo de esta teoría.

El cerebelo es una región importante del sistema nervioso responsable de la función motora, ubicada en la parte posterior de nuestra cabeza. A pesar de pesar solo 150 gramos y constituir aproximadamente el 10% del volumen total del cerebro, tiene cuatro veces más neuronas que todas las demás regiones del cerebro juntas y está altamente conectado con este.⁶⁸

Ahora bien, a pesar de esto, si este órgano fuera eliminado del lector, no tendría que preocuparse demasiado. Su conciencia seguiría completamente intacta. ¿Cómo puede ser esto posible? Según la TII, debido a que el cerebelo está compuesto por módulos independientes entre sí, la información procesada allí no está tan integrada como en otras áreas del cerebro, por lo tanto, su Φ es inferior al mínimo necesario para que haya conciencia.

¿Recuerdas algunas ideas que avanzamos en la primera parte sobre el sueño? La TII también predice algo interesante sobre lo que sucede en nuestro cerebro y nuestra conciencia cuando dormimos. Suponiendo entonces que la conciencia puede medirse en Φ , es intuitivo suponer que cuando dormimos sin ninguna experiencia consciente —por ejemplo, sin entrar en sueños lúcidos— nuestro cerebro deberá tener una

⁶⁸ Publicación original: Lemon, R. & Edgley, S. (2010) "Life without a cerebellum", *Brain*, 133: 652–654.

disminución significativa en la conectividad neuronal y, por lo tanto, estará menos integrado entre sí.

Ahora, bien, otros estudios empíricos en cerebros humanos durante el sueño muestran que esta predicción es efectivamente precisa. El cerebro disminuye su conectividad cortical.⁶⁹

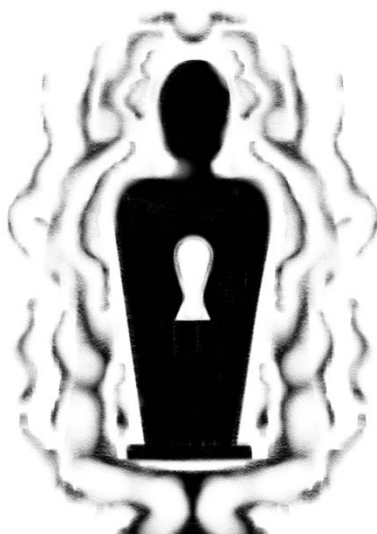


Imagen 23: "Il tassello mancante" © Gaia Amoroso

Otra predicción intrigante de esta teoría está relacionada con un tipo de procedimiento neuroquirúrgico conocido como "calosotomía" (en

⁶⁹ Publicación original: Massimini, M., ..., Tononi, G. (2005) "Breakdown of cortical effective connectivity during sleep", *Science*, 309: 2228-2232.

inglés, split-brain). Este procedimiento realiza algo curioso, divide el cerebro en dos hemisferios desconectados entre sí mediante una incisión precisa en el cuerpo calloso.

En un famoso experimento científico con pacientes que sufren de epilepsia,⁷⁰ después de la remoción del cuerpo calloso -para reducir los síntomas-, algo que predice la TII ocurre.

Si uno de estos pacientes visualiza una imagen solo con el ojo derecho, procesando la información visual a través del hemisferio izquierdo del cerebro y se le pregunta qué veía en ese momento, dirá que no puede recordar haber visto ninguna imagen.

Sin embargo, si pedimos al mismo paciente que dibuje lo que vio, ahora será capaz de dibujar correctamente lo que había observado. ¿Cómo puede ser esto posible? ¿Estamos ante algún tipo de magia demoníaca?

No obstante, si pedimos al mismo paciente que dibuje lo que vio, ahora será capaz de dibujar correctamente lo que observó. ¿Cómo puede ser esto posible? ¿Estamos frente a algún tipo de magia demoníaca?

Claro que no. La razón de este fenómeno se debe al hecho de que diferentes funciones están localizadas en diferentes partes del cerebro. Mientras que el lenguaje

⁷⁰ Publicación original: Gazzaniga, M. (1967) "The split brain in man", *Scientific American*, 217 (2): 24-29.

es una función cognitiva presente principalmente en el hemisferio izquierdo (que no procesó ninguna información que estaba siendo asimilada en el hemisferio derecho, dado que la conexión entre ambos estaba interrumpida), la función relacionada con el dibujo está presente en el hemisferio derecho.

Ahora bien, según la TII, cuando el cuerpo calloso es desconectado, el cerebro deja de ser, en su totalidad, un sistema completo, creándose dos sistemas independientes en cada hemisferio que pueden maximizar su Φ y, por lo tanto, generar dos corrientes de conciencia independientes entre sí.

Otro ejemplo relacionado con el mal funcionamiento del cerebro se encuentra en las lesiones cerebrales: esta teoría predice que cualquier tipo de lesión que afecte la capacidad del cerebro para integrar información impedirá la formación de estados conscientes.

Estudios recientes muestran que esta predicción es exactamente fiel a la realidad: mediante la estimulación magnética transcraneal, varios pacientes con lesiones cerebrales mostraron alteraciones sustanciales en sus cerebros en comparación con pacientes sanos.⁷¹

⁷¹ Publicación Original: Casali, A. et al. (2013) "A theoretically based index of consciousness independent of sensory processing and behavior", *Science Translational Medicine*, 5: 198ra105–198ra105.

¿Puede esta teoría, además de proporcionar una narrativa explicativa de cómo el cerebro biológico es consciente, arrojar algo de luz sobre el desarrollo de sistemas artificiales conscientes? La respuesta es mixta: por un lado sí, y por otro, no.

En el lado positivo, en teoría, la TII no especifica que un sistema tenga que ser biológico para ser consciente, solo indica que debe ser un sistema que pueda integrar información siguiendo los axiomas mencionados anteriormente.

Sin embargo, en el lado negativo, para los defensores de esta teoría, está claro que la inteligencia artificial actual, basada en estructuras de alimentación directa (feed-forward) en las que la información solo se procesa en una dirección (de entrada a salida), nunca podrá ser consciente, dado que un sistema de este tipo no tendrá ningún Φ , incluso si es un sistema que replique el comportamiento consciente a la perfección: eso sería solo una simulación. En su lugar, sería necesario que el sistema tuviera una estructura de retroalimentación (re-entry) en la que las salidas puedan ser utilizadas como entradas y viceversa.

Para concluir estas notas introductorias sobre la TII, falta abordar una crítica —que puede ser vista como una ventaja— planteada a esta teoría. Para algunos autores, la TII puede tener consecuencias pampsiquistas contraintuitivas. ¿Y qué es el pampsiquismo? El pampsiquismo es una teoría

desarrollada por varios pensadores a lo largo de la historia —desde Spinoza hasta Leibniz, Bertrand Russell y, más recientemente, Philip Goff⁷²— que sostiene que todo lo que existe en el mundo físico posee propiedades mentales fundamentales.

Esto significa afirmar que la conciencia es una propiedad fundamental, lo cual concuerda con la idea planteada por la TII de que algunos sistemas complejos, cuando alcanzan altos niveles de información integrada, se vuelven conscientes.

Además, tanto el pampsiquismo como la TII coinciden en que la conciencia no es un fenómeno de todo o nada, por el contrario, la conciencia debe ser vista como un espectro, entre lo absolutamente inconsciente y lo completamente consciente.

A pesar de estas similitudes, parece algo radical identificar la TII como una teoría pampsiquista. Aunque ambas tienen compromisos cartesianos al aceptar la conciencia subjetiva como fundamental en su naturaleza, el pampsiquismo sostiene que la conciencia es fundamental para todas las entidades existentes, mientras que la TII solo sugiere las condiciones

⁷² Publicación original: Goff, P. (2019) *Galileo's Error: Foundations for a New Science of Consciousness*, London: Pantheon Book.

necesarias para que la conciencia pueda emerger en determinados sistemas complejos.⁷³

Para finalizar estas notas introductorias, abordaremos otra perspectiva de mirar el cerebro a través de las emociones y el papel de estas en el desarrollo de nuestra mente, presentando algunas de las tesis centrales desarrolladas por el neurocientífico y director del “Emotional Brain Institute” en la Universidad de Nueva York, el profesor Joseph LeDoux.

⁷³ Sección basada en: Tononi, G. (2012) “Integrated information theory of consciousness: An updated account”, *Archives italiennes de biologie*, 150: 56-90; Tononi, G. & Koch C. (2015) “Consciousness: Here, There and Everywhere?”, *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 370 (1668): 20140167.

V. EL CEREBRO EMOCIONAL

En este cuarto y último diálogo de esta odisea por el mundo del cerebro consciente, presentaremos algunos de los conceptos clave desarrollados por el neurocientífico y autor del libro "The Emotional Brain", el profesor Joseph LeDoux. El primer paso para entender qué es este cerebro emocional es comprender cuáles son las estructuras anatómicas que componen este sistema vital para la supervivencia de cada uno de nosotros.

En neurociencia, la parte del cerebro que procesa las diversas emociones se denomina "sistema límbico", que está compuesto por varias partes y se encuentra en la superficie medial del cerebro de los mamíferos. Este sistema desempeña un papel fundamental en el funcionamiento de nuestras emociones y por consiguiente, en la forma en que nos relacionamos con nosotros mismos y con los demás.

A través del sistema nervioso autónomo, influye valorativamente (de forma positiva o negativa) en el funcionamiento visceral y la regulación metabólica del organismo, dirigiendo los comportamientos que se consideran esenciales para la supervivencia de todos los mamíferos.

La respuesta emocional apropiada se producirá entonces a través de la integración de la información sensorial en un estado mental particular, asignando un contenido afectivo a los estímulos que se registran y se asocian a memorias ya existentes.

Veamos algunas de las principales estructuras anatómicas que pertenecen al sistema límbico:

- **Amígdala:** Localizada en el lóbulo temporal anterior, está conectada al hipotálamo y se considera el "centro" que nos orienta en situaciones de peligro, produciendo estados de alerta o sensaciones de miedo. Su eliminación en mamíferos causa una postura sexualmente indiscriminada, indiferencia a situaciones de riesgo y pérdida del sentido afectivo de los estímulos exteriores. Por otro lado, su estimulación eléctrica genera comportamientos agresivos y violentos.
- **Hipotálamo:** Localizado en el centro del cerebro, justo debajo del tálamo y encima de la hipófisis, es considerado la parte más fundamental del sistema límbico, ya que es responsable de las funciones vegetativas del cerebro (mantenimiento del equilibrio interno, temperatura corporal, hambre, sed, producción hormonal, etc.). Además, está directamente relacionado con el comportamiento, desempeñando un papel esencial en las

emociones (la parte intermedia está ligada a la aversión y las partes laterales al placer y la ira).

- Tálamo: Localizado en el diencéfalo, justo debajo del cuerpo caloso, es la parte del sistema límbico donde se selecciona la información sensorial captada del ambiente. También es responsable de la parte motora (activación muscular) y está relacionado con los estados de alerta. Se considera fundamental en la percepción del dolor y es importante en la regulación de los estados de sueño y vigilia.
- Hipocampo: localizado en la parte interna del lóbulo temporal medial, es responsable de la memoria a corto y largo plazo, permitiendo al organismo comparar situaciones actuales con experiencias pasadas, aumentando así su capacidad de supervivencia. También es responsable de la orientación y navegación espacial.

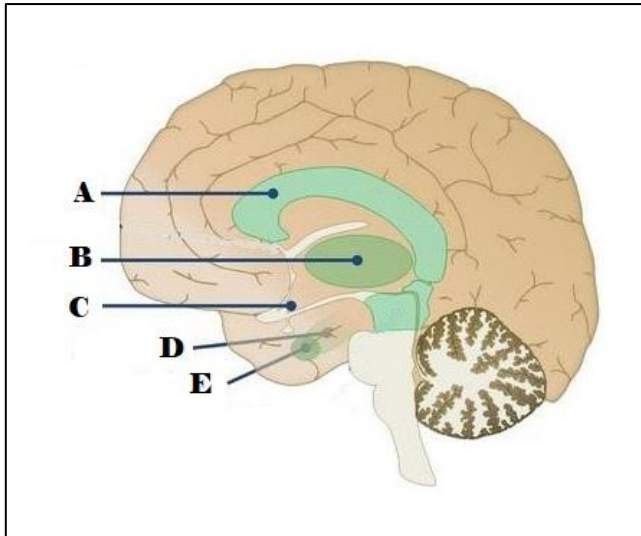


Imagen 24: Sistema Límbico: A = Cuerpo calloso; B = Tálamo; C = Hipotálamo; D = Hipocampo; E = Amígdala.

Partiendo de esta breve descripción neuroanatómica, podemos comprender mejor las diversas tesis avanzadas por Joseph LeDoux. Este neurocientífico propone una nueva conceptualización que redefine algunos conceptos clave en el estudio de las emociones, especialmente las funciones y circuitos relevantes para la supervivencia de los organismos.

En este nuevo enfoque, se busca revertir el foco de la forma tradicional de investigar las emociones, en lugar de cuestionarnos si las emociones sentidas conscientemente por los seres humanos están también presentes en diversos animales – en una actitud clara

de antropomorfismo –, se intenta plantear la cuestión de comprender hasta qué punto las funciones y circuitos relevantes para la supervivencia y preservación de los animales están, o no, también presentes en nosotros, los seres humanos.

Es importante comprender que, para LeDoux, estos circuitos y funciones no están relacionados causalmente con las emociones, aunque contribuyen indirectamente a ellas. ¿Qué son, entonces, estos circuitos de supervivencia? En el fondo, este concepto amplio busca integrar todo tipo de mecanismos relevantes para la preservación y el éxito del organismo en su vida cotidiana.

¿Y por qué introducir este nuevo concepto en lugar de hablar de “emociones” directamente? LeDoux propone un argumento filosófico muy interesante para apoyar este cambio. Para él, el concepto de “emoción” es filosóficamente inconsistente, no habiendo consenso en la comunidad científica sobre su significado real. Por eso sugiere esta inversión metodológica, para no quedar “atrapado” en esta indefinición conceptual.

Siguiendo esta interesante innovación filosófica, LeDoux buscará centrarse en los fenómenos que son relevantes en el estudio de las emociones, evitando, sin embargo, el uso de conceptos emocionales particulares. En lugar de eso, el enfoque estará en los circuitos específicos que instancian determinadas funciones que permiten al organismo sobrevivir a

través, por ejemplo, de la detección y respuesta a desafíos y oportunidades. Serán estas instancias particulares las que podrán ser etiquetadas con alguna emoción al final (pero no viceversa).

Esos circuitos implican desarrollar capacidades para identificar el peligro, posibles parejas de apareamiento, la presencia de comida o agua, la defensa y el mantenimiento energético, la termorregulación, entre muchos otros. En el fondo, el desafío del científico será describir todos estos procesos evitando caer en el uso de un lenguaje confuso y ambiguo, como es el lenguaje de los conceptos relacionados con las emociones.

Observe que estos circuitos de supervivencia se conservan en todas las especies de mamíferos (al menos, y quizás en muchos otros animales y organismos). Aunque puedan existir algunas diferencias, hay componentes esenciales de esas funciones que son compartidos por todos los animales.

Vemos así de qué forma la inversión ledouxiana es útil: dado que esas funciones de supervivencia están instanciadas en circuitos conservados por la historia evolutiva, podemos evitar antropomorfizar al cuestionarnos "¿Qué emociones humanas están presentes en los otros animales?" y preguntar, en cambio, "¿Qué circuitos presentes en los otros animales están presentes en los seres humanos?"

La idea de esta inversión es pensar en las emociones y conceptos adyacentes (como la motivación, el refuerzo, la inhibición, la excitación, etc.) como componentes de un proceso mayor, que no se agota en un sentimiento específico.

Por lo tanto, lo que se sigue de este enfoque no es un intento de explicar o definir las emociones. Por el contrario, el objetivo es proporcionar una estructura para pensar algunos fenómenos asociados a las emociones – fenómenos relacionados con la supervivencia – de una manera que no se confunda con la búsqueda del significado demasiado amplio de "emoción" y permitiendo el enfoque en aspectos clave que eviten debates infinitos sobre una definición correcta de "tristeza", "miedo", "felicidad", entre todos los otros conceptos que usamos para describir nuestros sentimientos ("feelings").

Este enfoque sobre la relación de las funciones de supervivencia conservadas por la evolución con las emociones no es completamente nuevo, siguiendo, por el contrario, una tradición que se remonta al menos a Darwin (1872). No será una innovación para los neurocientíficos que estos circuitos estén de hecho vinculados al cerebro.

La estrategia e innovación de LeDoux es centrarse precisamente en estos circuitos para sustancializar las diferentes emociones, en lugar de la estrategia contraria típicamente utilizada hasta entonces, de

partir de definiciones a priori de emociones y solo después intentar identificar los circuitos relacionados con las mismas. Para simplificar, el objetivo es identificar primero el circuito y solo después la emoción, en vez de comenzar por la emoción para identificar el circuito.

Una segunda problemática interesante en la investigación de las emociones es comprender si existen circuitos emocionales en el cerebro que son innatos en vez de creaciones de la mente humana o de los aspectos sociales y culturales. La complejidad de este debate radica en la cuestión de entender si las llamadas "emociones básicas" son o no naturales.

Las emociones básicas son aquellas que, supuestamente, son expresadas por todas las personas que existen, estando también presentes en los animales más cercanos, en términos evolutivos, a los seres humanos. El ejemplo canónico de emociones básicas es la lista de Ekman (1972, 1992)⁷⁴ que incluye la felicidad, el miedo, la tristeza, la ira, la sorpresa y el asco, siendo que muchas investigaciones de la

⁷⁴ Publicación original: Ekman, P. (1972) "Universals and cultural differences in facial expression of emotion" In J. Cole (Ed.), *Nebraska Symposium on Motivation*, Lincoln, Nebraska: University of Nebraska Press: pp. 207-283. An interesting update can be found at: Ekman, Paul (1992). "An Argument for Basic Emotions", *Cognition and Emotion*, 6 (3/4): 169-200.

neurociencia actual buscan identificar la base neuronal de esas emociones.

A pesar de estas múltiples investigaciones neurocientíficas, el concepto de "emociones básicas" ha planteado varios desafíos, dado que existen diversas teorías que incluyen otras emociones o excluyen algunas de la lista de Ekman. Claro, esto probablemente se deba a la dificultad de etiquetar un concepto emocional con una diversidad de estados biológicos que pueden o no coincidir con ese concepto.



Imagen 25: "Nosce te ipsum" © Gaia Amoroso

Además, hay dos dificultades centrales más en este enfoque de las "emociones básicas": (i) por un lado,

algunos sostienen que las emociones son construcciones psicológicas y sociales que surgen de la interacción con el entorno físico o social, sugiriendo que no hay una determinación biológica previa; (ii) la mayoría de las teorías sobre las emociones básicas se basan en la investigación de cerebros de animales que no encajan del todo en la lista de Ekman u otras emociones básicas.⁷⁵

Por ejemplo, Louise Barrett considera que no tiene sentido argumentar que las emociones básicas han sido conservadas por la evolución a través de circuitos neuronales por tres razones particulares⁷⁶:

(i) algunos sostienen que las emociones son construcciones psicológicas y sociales que surgen de la interacción entre el entorno físico o social; por lo tanto, no existe una determinación biológica inherente desde el principio;

(ii) la gran mayoría de las teorías asociadas con las emociones básicas se basan en investigaciones sobre cerebros de animales que no se alinean con la lista de Ekman ni con otras

⁷⁵ Publicación original: Panksepp, J. (1998) *Affective Neuroscience*, New York: Oxford U. Press. Uma atualização relevante poder ser encontrada em: Panksepp J. (2005) "Affective consciousness: Core emotional feelings in animals and humans", *Consciousness and Cognition*, 14: 30–80.

⁷⁶ Publicación original: Barrett, L. (2006) "Are Emotions Natural Kinds?", *Perspectives on Psychological Science*, 1:28–58.

emociones básicas propuestas (Panksepp, 1998, 2005).⁷⁷

Por ejemplo, Louise Barrett (2006)⁷⁸ considera que no tiene sentido argumentar que las emociones básicas se conservaron a través de la evolución a través de los circuitos neuronales por tres razones particulares:

- Algunos estudios de imágenes muestran que áreas activadas por el mismo estímulo han sido asociadas a diferentes emociones básicas;
- Gran parte de las evidencias a favor de las emociones básicas se han recogido en cerebros de animales (y no humanos) mediante técnicas retrogradas (por ejemplo, estimulación eléctrica del cerebro) con una falta de precisión notable.
- Las emociones básicas identificadas en estudios en seres humanos no coinciden con las emociones básicas identificadas en animales.

Ahora bien, esta controversia podría aclararse nuevamente mediante el enfoque de LeDoux. En lugar de hablar de "emociones básicas", los científicos y filósofos deberían optar por "circuitos de

⁷⁷ Publicación original: Panksepp, J. (1998) *Affective Neuroscience*, New York: Oxford U. Press. A relevant update can be found at: Panksepp J. (2005) "Affective consciousness: Core emotional feelings in animals and humans", *Consciousness and Cognition*, 14: 30–80.

⁷⁸ Publicación original: Barrett, L. (2006) "Are Emotions Natural Kinds?", *Perspectives on Psychological Science*, 1:28–58.

supervivencia". Esto les permitiría resolver estas y otras confusiones conceptuales que surgen del uso del lenguaje emocional en primer lugar y que tienen implicaciones directas en la investigación empírica.

Estos circuitos de supervivencia incluyen circuitos relacionados con la nutrición, los fluidos, la defensa, la reproducción y la termorregulación. Estos circuitos ya estarían presentes en las primeras formas de vida en la tierra; por ejemplo, las bacterias pueden escapar de entornos tóxicos y buscar entornos favorables para su desarrollo utilizando algunos de estos circuitos.

Con el desarrollo de organismos eucarióticos multicelulares, estos circuitos de supervivencia se volvieron cada vez más complejos, especialmente con el desarrollo de receptores sensoriales especializados y un sistema nervioso central capaz de coordinar todas las funciones corporales y diversas interacciones con el ambiente.⁷⁹

Debe quedar claro, sin embargo, que el objetivo de LeDoux al describir los diversos circuitos de supervivencia no es vincularlos directamente con las

⁷⁹ Publicación original: Shepherd, G. (1988) *Neurobiology II*, New York: Oxford. Así que, contrariamente a mucha investigación en neurociencia, debe notarse que un circuito de supervivencia no está directamente vinculado a una emoción particular (mucho menos a una activación neuronal específica en el cerebro). Por el contrario, diferentes circuitos de supervivencia pueden estar relacionados con una misma emoción y viceversa.

"emociones básicas". Más bien, el objetivo es liberar el lenguaje neurocientífico de conceptos basados en sentimientos subjetivos del ser humano y enfocarse de manera neutral en esos circuitos conservados a lo largo de la historia evolutiva.

Para concluir, veamos un breve ejemplo en relación con la agresión para entender la relevancia de este enfoque. Sabemos que la agresión es una emoción que no tiene una única "representación" fija en el cerebro; más bien, existen diferentes tipos de agresión dependiendo del contexto en el que ocurre.

Por ejemplo, la agresión puede ser desencadenada por el circuito de defensa (en la protección del organismo); puede ser desencadenada por el circuito alimentario (en la caza de alimentos); o puede ser desencadenada por el circuito reproductivo (en la competencia por parejas).

Así que, contrariamente a mucha investigación en neurociencia, debe notarse que un circuito de supervivencia no está directamente vinculado a una emoción particular (mucho menos a una activación neuronal específica en el cerebro). Por el contrario, diferentes circuitos de supervivencia pueden estar relacionados con una misma emoción y viceversa.

Esto podría significar que estos circuitos no tienen una relación directa ni un papel causal con las emociones. Claro, seguramente haya una influencia indirecta, pero

el enfoque principal de estos circuitos es garantizar propósitos adaptativos para asegurar la preservación del organismo frente a las diversas complejidades requeridas por la interacción con el ambiente.

Por lo tanto, ahora queda claro de qué manera la inversión propuesta por LeDoux es útil: si examinamos los circuitos de supervivencia compartidos por animales y seres humanos, y solo después consideramos las llamadas "emociones", estaremos más cerca de asegurar explicaciones relevantes y comprender la naturaleza y el papel de esas emociones en el ser humano.⁸⁰

Concluidas estas segundas notas introductorias, el lector podrá disfrutar de los próximos diálogos con otros cuatro pensadores de renombre internacional sobre la naturaleza del cerebro desde múltiples perspectivas, lo cual sin duda enriquecerá su conocimiento sobre esta fascinante odisea que ocurre dentro de su cabeza.

⁸⁰ Sección basada en: LeDoux, J. (1996) *The Emotional Brain*, New York: Simon and Schuster; LeDoux, J. (2003) "The Emotional Brain, Fear and the Amygdala", *Cellular and Molecular Neurobiology*, 23 (4/5): 727-738; LeDoux, J. (2012) "Rethinking the emotional brain", *Neuron*, 73 (4): 653-676.

DIÁLOGOS II

Cerebro

DIÁLOGOS II

Cerebro

VI. Diálogo con Anil Seth



Anil Seth es profesor de "Neurociencia Cognitiva y Computacional" en la Universidad de Sussex, donde también es co-director del "Sackler Center for Consciousness Science". Además, es co-director del programa del "Instituto Canadiense de Investigación Avanzada" (CIFAR) sobre "Cerebro, Mente y Conciencia", así como del programa de becas de doctorado Leverhulme titulado "De la Sensación y Percepción a la Conciencia".

Es el editor de la revista "Neuroscience of Consciousness" (Oxford University Press). Fue el presidente de la conferencia para el "16º Encuentro de la Asociación para el Estudio Científico de la Conciencia" (ASSC16, 2012).

Su investigación ha sido financiada por EPSRC (Leadership Fellowship), el "Consejo Europeo de Investigación" (ERC, Advanced Investigator Grant), el "Wellcome Trust" y el "Instituto Canadiense de Investigación Avanzada" (CIFAR).

Ha publicado varios libros, siendo el más reciente "Being You: a New Science of Consciousness" (Faber & Faber), que fue considerado uno de los 7 principales libros de ciencia por el periódico "The Guardian" en 2021.

Para más información: <https://www.anilseth.com/>

Pregunta: El Profesor Seth es uno de los principales neurocientíficos que defiende la “Teoría del Procesamiento Predictivo”, una idea con una larga historia y con supuestos interesantes sobre el cerebro y la mente humana. ¿Cuándo se encontró el Profesor Seth por primera vez con la idea de que el cerebro es una máquina de predicciones?

Anil Seth: La primera vez que escuché sobre esta idea, me pareció de inmediato una manera realmente atractiva de pensar no solo en cómo funciona el cerebro en general, sino también en los temas que me interesaban particularmente en ese momento y aún me interesan. Por ejemplo, cómo pensamos en los correlatos neuronales de la conciencia o cómo relacionamos lo que sucede en el cerebro con nuestras experiencias conscientes, así como también cómo comprendemos las emociones en el “yo”.

Pensar en estas ideas de “codificación predictiva”, de las cuales probablemente escuché por primera vez hace unos veinte años, y en las cuales comencé a trabajar hace unos doce o trece años, parecía ser el tipo de lenguaje relevante para utilizar en la investigación.

Otro aspecto que realmente me atrajo fue que la codificación predictiva es algo que se puede abordar en varios niveles o disciplinas; a un nivel muy matemático, a un nivel computacional (se puede escribir código predictivo), pero también a un nivel conceptual.

Además, también podemos abordarla a nivel filosófico, lo que la convierte en una excelente forma de unir estas diversas disciplinas. Creo que esta característica multidisciplinaria es muy relevante en la neurociencia actual.

Pregunta: ¿Podría contarnos brevemente la historia de esta perspectiva? ¿Quiénes son los primeros o principales defensores y proponentes de la idea de que el cerebro es un agente activo en el mundo?

Anil Seth: Platón (risas)! Claro, depende de hasta qué punto queramos remontarnos en la historia. Sabemos que ya en la filosofía griega podemos encontrar la idea de que lo que consideramos real, lo que percibimos en los contenidos de nuestra experiencia, no necesariamente es lo que realmente existe.

Es una reflexión similar a la que Platón tiene en la “Alegoría de la Caverna”, donde habla de los prisioneros encadenados a la pared de la caverna que consideran las sombras en la caverna como el mundo real. Ellos no conocen y no pueden conocer ninguna diferencia más allá de la realidad y la ilusión de la caverna. Sin embargo, claro está, aquí no hay una idea explícita de codificación predictiva como principio de la función cerebral.

También existe el concepto hindú de “Maya”, que enfatiza el proceso mediante el cual lo que percibimos

tiene el carácter de parecer real. Esto está también íntimamente relacionado con los conceptos modernos de procesamiento predictivo.

En el siglo XV, encontramos a Ibn al-Haytham, un estudioso árabe que probablemente fue la primera persona en pensar en la percepción como inferencia, es decir, la idea de que lo que percibimos es un juicio sobre lo que está ocurriendo en el mundo, algo que el organismo estima o predice.

Otra gran influencia es el científico alemán del siglo XIX Hermann von Helmholtz, quien abordó esta perspectiva con mucho más detalle, presentando la primera formulación de la idea de que la percepción es un proceso de inferencia basado en el cerebro que es en sí mismo inconsciente. La formulación de von Helmholtz se alineó naturalmente con las matemáticas de la inferencia proveniente del razonamiento bayesiano.

Pero la codificación predictiva en sí misma, como idea y algoritmo específico, surgió de la ingeniería y el procesamiento de señales, donde los científicos intentaban descubrir formas de comprimir señales largas sin perder información. La idea es: si podemos prever algo, entonces no necesitamos codificarlo explícitamente.

Siguiendo esta idea, el algoritmo de codificación predictiva fue desarrollado en la década de 1950 en el

campo de la ingeniería. Alrededor de 1990, se publicó el primer artículo o referencia clásica centrada en el sistema visual. Rajesh Rao y Dana Ballard defendieron la tesis de que el sistema visual puede estar incorporando un tipo de inferencia predictiva y utilizaron esta perspectiva para explicar las propiedades del sistema visual.

Desde entonces, el procesamiento predictivo ha sido considerado una teoría mucho más general sobre cómo funciona el cerebro y es utilizada por muchos pensadores y científicos para elucidar diversos aspectos del ser humano.

Pregunta: El profesor Seth adopta una posición crítica respecto a la famosa distinción introducida por David Chalmers entre los problemas fáciles (easy) y difíciles (hard) de la conciencia. En el fondo, no acepta esta división y considera más interesante hablar del problema real de la conciencia. ¿Podría clarificar un poco más sobre esta estrategia de cambio conceptual para abordar la conciencia desde un punto de vista científico y objetivo?

Anil Seth: Le llamo el problema real de la conciencia para "fastidiar" a David Chalmers, dado que él ha contribuido tanto y de manera tan significativa a la filosofía de la mente y considero que ha articulado muy bien esa distinción entre problemas difíciles y fáciles.

Sin embargo, creo que dicha distinción ha hecho que el problema de la conciencia sea un tanto obstruido. Si aceptamos esta división, estamos aceptando la intuición de que ninguna explicación del cerebro o de cualquier sistema físico podría explicar qué partes del universo son conscientes y qué otras partes no lo son.

Es cierto que la conciencia no parece ser el tipo de cosa que podría explicarse en términos físicos, lo cual es una intuición realmente profunda. Y, desde cierta perspectiva, esta intuición es precisamente lo que el problema difícil de la conciencia formaliza.

Pero si aceptamos esta dificultad, surge un efecto secundario, terminamos tratando la conciencia como un gran misterio y tratamos de encontrar una solución "mágica", una fuente especial, quizás algún tipo de misterio cuántico, del cual esperamos que pueda surgir la magia de la experiencia consciente a partir de un simple mecanismo.

Y tal vez sea necesario algo mágico, en el sentido de que podríamos necesitar una nueva física o un nuevo momento "eureka". Sin embargo, creo que un enfoque más interesante es centrarnos en el problema real de la conciencia. Nuevamente, esta no es una idea nueva, simplemente es una manera de enfatizar más dentro del lenguaje que Chalmers ha hecho tan popular.

El problema real de la conciencia es la idea de afirmar: "Bien, puede que no sepamos cómo los sistemas físicos

generan la conciencia o son idénticos a ella, pero sabemos que la conciencia existe y que está íntimamente relacionada, de varias formas, con el cerebro. La conciencia es un hecho en el universo en el que vivimos y que necesita ser explicado."

La aproximación al problema real argumenta que no deberíamos abordar el problema "de frente", como intentan hacer las soluciones al problema difícil, sino que podemos dividir un poco el problema en partes y tratar de entender cuáles son las propiedades de la conciencia y cómo podemos explicar esas propiedades.

La analogía que utilizo en mi reciente libro "Being You" (2022) es la idea de que los científicos, no hace mucho tiempo, pensaban que la vida no podía ser explicada en términos de mecanismo. Debido a esta idea, los pensadores de la época pasaron mucho tiempo discutiendo si había una centella de vida, un "Élan Vital" u alguna otra cosa.

Pero al final, los científicos no encontraron esa chispa mágica de vida y tampoco llegaron a la conclusión de que la vida no existía. Lo que ocurrió fue que simplemente se dieron cuenta de que la vida no es algo único o unitario, sino que está compuesta por muchas propiedades diferentes y que podemos explicar cada una de esas partes mediante la química y la física.

Siguiendo esta estrategia, el problema difícil de la vida no fue resuelto, sino más bien disuelto: es este tipo de

enfoque el que creo que puede ser muy útil para el estudio de la naturaleza de la conciencia, aunque eventualmente no tenga éxito en disolver completamente el problema difícil. Pero solo sabremos si funciona o no si lo intentamos y este enfoque real busca hacer precisamente eso, al menos intentarlo.

Pregunta: Por lo tanto, este enfoque o cambio conceptual podría, al final, incluso crear nuevos problemas que inicialmente no podrían habernos surgido, al disolver esa distinción, algo problemática, entre el problema fácil y difícil de la conciencia, podríamos abrir una nueva línea de investigación que podría conducir a algún progreso. Antes de esto, si afirmamos que la conciencia es difícil en el sentido de Chalmers, entonces no podríamos avanzar mucho en términos de intentar desarrollar investigaciones empíricas sobre la naturaleza de la conciencia.

Anil Seth: Exactamente. Creo que este es un punto muy relevante, porque realmente comprende por qué este cambio conceptual es pertinente. Creo que es un aspecto muy subestimado de la interacción entre ciencia y filosofía. No es que tengamos este "menú" fijo de preguntas y solo podamos resolverlas una por una, aunque a menudo se presenta así al público.

Es tan común ver artículos hoy en día, tanto en periódicos como en revistas científicas, que tratan

sobre los "10 mayores misterios no resueltos". Sin embargo, creo que este enfoque descuida por completo la dialéctica entre teorías, datos y preguntas, es posible avanzar en el conocimiento del mundo tanto cambiando las preguntas que formulamos como desarrollando nuestras respuestas.

Pregunta: Y podemos incluso descubrir que la pregunta o el problema principal que estábamos tratando de resolver inicialmente no tiene sentido después de todo, ¿verdad?

Anil Seth: Claro. Volviendo a la analogía anterior, si alguien hoy en día intenta obtener una beca de investigación para buscar el "Elan Vital", seguramente no la conseguirá. Es una pregunta que no tiene sentido alguno. Lo mismo ocurrió en la física con el antiguo concepto de "éter".

¿Qué implica eso? Creo que implica que el aparato conceptual que heredamos de problemas o misterios específicos siempre debe ser cuestionado, ya que está siempre disponible para nosotros, lo que podría llevarnos a utilizarlo de manera acrítica y no ver los problemas que esta aceptación neutra puede plantear para nuestra propia investigación.

Pregunta: Volviendo a la “Teoría del Procesamiento Predictivo”, el profesor Seth cree que este enfoque predictivo hacia el cerebro y la mente podría arrojar algo de luz sobre el desarrollo de una “Inteligencia Artificial” consciente. Escuchamos mucho sobre la “conciencia artificial” en estos días, pero parece no tener mucho sentido intentar desarrollar una conciencia “artificial” sin primero comprender cómo existe la conciencia “natural”.

Anil Seth: Para empezar, creo que ya existe una gran contribución del “Procesamiento Predictivo”, como se considera en la Ciencia Cognitiva, en la Inteligencia Artificial. Como mencioné antes, fue en la ingeniería donde surgió el procesamiento predictivo como método de compresión de señal.

Además, podemos encontrar objetos interesantes en Inteligencia Artificial como las máquinas de Helmholtz, en honor al mismo Helmholtz que originó el concepto de percepción como inferencia inconsciente, y codificadores automáticos, e incluso algoritmos que implementan efectivamente aspectos de lo que podríamos llamar teoría del procesamiento predictivo en las ciencias cognitivas: las principales computaciones que constituyen esa teoría.

Este tipo de sistemas ya difieren un poco de las redes de aprendizaje profundo (deep learning) que son tan predominantes en la investigación y en las aplicaciones de Inteligencia Artificial actualmente.

Los principales ingredientes para el procesamiento predictivo e Inteligencia Artificial son los siguientes: Necesitamos modelos generativos, algo que sea capaz de generar previsiones sobre sus señales de entrada (“inputs”). A continuación, necesitamos un modelo que compare esas señales de entrada con esas previsiones y actualice las previsiones anteriores y, a largo plazo, el propio modelo. A partir de aquí, podremos crear formas de percepción y de aprendizaje con el mismo algoritmo.

De hecho, uno de mis colegas, un talentoso ex-estudiante de doctorado llamado Beren Millidge, demostró que, si implementamos la codificación predictiva, realmente podemos obtener retropropagación (back propagation), que es el tipo de algoritmo de aprendizaje más usado en Inteligencia Artificial.

Muchos investigadores han criticado la retropropagación porque afirmaban que no sería neurobiológicamente plausible, no sería el tipo de cosa que hacen los cerebros, dado que requiere señales de error globales que se propagan por toda la red.

Sin embargo, lo que descubrimos fue que la codificación predictiva proporciona la retropropagación de forma gratuita, de una manera que es (potencialmente) biológicamente plausible. Esto representa una evidencia más fuerte de que nuestros cerebros realmente funcionan así.

En la actualidad, mucha investigación en Inteligencia Artificial está siguiendo ese camino y construyendo no solo redes cada vez más profundas, sino redes que poseen modelos generativos. Vemos este tipo de tecnología en todas partes, como las redes adversarias generativas (“adversarial networks”), algunas de las cuales son útiles para la neurociencia, mientras que otras pueden ser menos útiles. De cualquier manera, creo que hay un territorio realmente fértil en esa intersección.

Pregunta: El profesor Seth publicó su nuevo libro, “Being You: A New Science of Consciousness”, que fue muy bien recibido, siendo considerado por el “Financial Times” y “The Guardian” como el libro del año en la categoría de ciencia. ¿Qué idea principal del libro le gustaría destacar?

Anil Seth: Creo que hay tres “mensajes” principales que intenté transmitir en mi libro. El primero es que la conciencia puede ser abordada científica y filosóficamente sin enfrentar el problema difícil directamente, es decir, podemos tratar este problema real con un enfoque neurofenomenológico, y aquí debo dar crédito a Francisco Varela, quien fue el primero en proponer este enfoque para estudiar la conciencia.

La segunda idea es que podemos usar la codificación predictiva para intentar explicar no solo lo que hacen

los organismos -cómo reciben información sensorial para orientar el comportamiento-, sino también para mapear el contenido de las predicciones perceptivas. Desde esta perspectiva, diferentes tipos de experiencia visual pueden explicarse por involucrar diferentes tipos de predicción perceptiva.

La tercera y última idea es que todo este mecanismo predictivo también se aplica al “yo” (“self”), y este es el enfoque principal del libro. El “yo” no es la “cosa” que tiene percepciones, ni tenemos un pequeño “mini-Yo” dentro de nuestra cabeza que procesa todas las percepciones y decide qué hacer.

Lo que experimentamos como el “yo” también es una especie de inferencia perceptiva. En este caso, las señales sensoriales provienen en gran parte del interior del cuerpo, como la frecuencia cardíaca o el estómago, lo que podemos llamar interocepción.

Esta es la gran tesis de mi libro: todo lo que aparece en la conciencia es algún tipo de predicción perceptiva y todas las experiencias perceptivas que tenemos están, en última instancia, fundamentadas y basadas en el cuerpo y nuestra naturaleza como máquinas vivas. Existe una relación de conexión muy íntima entre el cuerpo vivido, el “yo” y la conciencia en general.

Pregunta: Esta aproximación puede dar forma o ser aplicada a muchas investigaciones en trastornos

psiquiátricos. ¿Podemos suponer que muchos de los trastornos mentales en cerebros "enfermos" pueden estar basados en algún tipo de desequilibrio entre la entrada predicha y la entrada real que debería ser procesada de alguna manera por el modelo inferido?

Anil Seth: Espero sinceramente que así sea el caso. Ciertamente es algo que hemos estado intentando abordar en mi centro de investigación en la Universidad de Sussex durante los últimos diez años. Creo que, con todos estos temas relacionados con la enfermedad mental y los trastornos psiquiátricos, hay muchas preocupaciones, pero también muchas promesas excesivas.

Considero que hay mucho por hacer, dado que en el tratamiento de la salud mental y en la medicina psiquiátrica, muchas de las aproximaciones todavía se centran principalmente en el tratamiento de los síntomas, en lugar de abordar las causas mecanicistas subyacentes.

Expandiendo un poco más sobre el tema, creo que la mayoría de los enfoques actuales abordan los síntomas de los trastornos mentales de la misma manera en que alguien toma paracetamol para aliviar los síntomas de una fiebre causada por una enfermedad subyacente.

Felizmente, para muchas enfermedades, ya tenemos tratamientos que abordan estas causas subyacentes, como los antibióticos, por ejemplo. Pero es cierto que,

en psiquiatría, en muchos casos, no tenemos realmente un equivalente. No existen "antibióticos psiquiátricos" que apunten a los mecanismos que causan los síntomas.

Estoy convencido de que intentar entender los trastornos psiquiátricos a través del prisma del procesamiento predictivo puede ofrecernos un camino mejor para identificar mecanismos plausibles.

Así mismo, todavía queda un largo camino por recorrer antes de que tengamos tratamientos efectivos en nuestras manos. Hay señales prometedoras, pero no existe una panacea que resuelva todo relacionado con los trastornos mentales. Ni siquiera los psicodélicos...

Pregunta: Para concluir, me gustaría saber tu opinión sobre el debate actual sobre el libre albedrío. Hay pensadores como Robert Sapolsky que argumentan que no existe ningún libre albedrío, abogando por un determinismo radical. Pero otros filósofos como Daniel Dennett sostienen que tenemos grados de libertad. ¿Cuál es tu posición en este debate?

Anil Seth: Tengo un capítulo en el libro exactamente sobre ese tema, que curiosamente es el único tema sobre el cual nunca había escrito antes.

El concepto de libre albedrío comenzó a tener más sentido para mí cuando empecé a pensar en él de la

misma manera que pienso sobre cualquier otro tipo de experiencia. Como una especie de predicción perceptual. Ahora, ¿Qué significa esto en términos de cómo se formula normalmente el debate? Hay personas que, como Sapolsky, defienden firmemente la tesis de que el libre albedrío no existe, y normalmente lo afirman debido a la sospecha de que el universo es de alguna manera determinista, es decir, que su curso está completamente definido desde el principio. Este es ciertamente el caso de Sapolsky.

Sin embargo, me parece que argumentos como este a menudo caen en la falacia del hombre de paja, argumentando en contra de una versión del libre albedrío que, de todas formas, no vale la pena mantener. Esta versión indeseable del libre albedrío es lo que los filósofos llamarían "libre albedrío libertario".

Soy muy escéptico de este tipo de libre albedrío y lo llamo "libre albedrío fantasmagórico" —que básicamente es la idea de que la conciencia puede activarse y alterar el curso de los eventos físicos, una especie de causa estratégicamente inteligente no causada. Este tipo de libre albedrío no tiene sentido, pero no creo que sea el tipo de libre albedrío que necesitamos o debemos defender.

El principal problema con las discusiones sobre el libre albedrío, desde mi punto de vista, es que las personas tienden a hacer suposiciones que no se dan cuenta de que están presuponiendo. Para mí, la pregunta que

realmente deberíamos hacer es la siguiente: ¿Es el cerebro un sistema lo suficientemente complejo como para reaccionar ante situaciones de manera muy flexible, que no están inmediatamente determinadas por el ambiente? Y la respuesta a esa pregunta, para organismos como nosotros, es: ¡sí!

Podemos, de hecho, realizar acciones cuyas causas provienen más desde dentro del cuerpo, y que se remontan profundamente a nuestras historias personales, más que del mundo exterior inmediato o del ambiente.

Temos muchos grados de libertad, nuevamente, en un sentido de ingeniería muy preciso. Puedo hacer muchas cosas, pero solo hago una de esas cosas. Y nuestro cerebro, altamente complejo, controla estos grados de libertad, integrando muchas causas anteriores de formas muy sofisticadas y sutiles en un único "camino final común" de comportamiento.

Paralelamente, es un hecho que en algunas ocasiones, cuando realizo ciertas acciones, experimento una sensación de voluntad y agencia. A veces, percibo una acción como voluntaria y otras veces no. Básicamente, adopto la misma perspectiva que con otros tipos de experiencias perceptivas y me pregunto: "¿Por qué esta experiencia parece ser así?" o "¿Cuál es el propósito de esta experiencia?".

Si seguimos esa línea de pensamiento, se puede entender las experiencias de libre albedrío como equivalentes a las experiencias de ver colores. No reflejan la realidad tal como es, pero aún así son muy útiles para el organismo. El color rojo no existe objetivamente como una propiedad del mundo, pero es algo muy útil de experimentar subjetivamente, ya que rastrea una propiedad útil de cómo se comportan los objetos en el mundo.

En mi opinión, las experiencias de libre albedrío rastrean una propiedad muy útil de cómo nosotros, como organismos complejos, nos comportamos. Estas experiencias etiquetan acciones que tienen una cierta "libertad del inmediatismo", para usar el término de Mike Shadlen. Y esto es útil, no porque las experiencias de libre albedrío causen realmente acciones, sino porque permiten que el organismo aprenda para que, la próxima vez, si las cosas no funcionan, se pueda tomar una acción diferente.

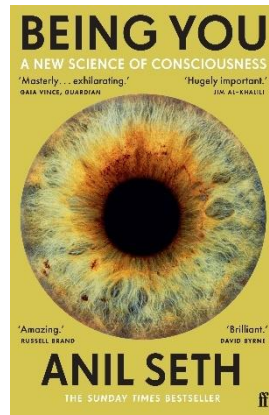
Desde mi punto de vista, el error es pensar que la experiencia del libre albedrío provoca una acción: eso no es lo que sucede. Nuevamente, la experiencia del libre albedrío es útil porque etiqueta acciones específicas como teniendo sus causas más internas que externas.

Se abordamos las cosas de esta manera, todo el debate sobre el determinismo y la estocasticidad se vuelve en gran medida irrelevante. Podemos pensar en el libre

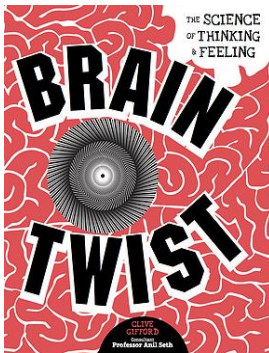
albedrío como un fenómeno biológico natural, pero es real en el mismo sentido en que experimentar los colores es real, y es útil en el mismo sentido en que la experiencia de los colores es útil.

Pensar en el libre albedrío de esta manera más sutil, como algo que existe en grados, significa que podemos hablar sobre los grados en los que varía, una discusión que se vuelve claramente más complicada cuando hablamos de responsabilidad moral y demás. Sin embargo, esta complejidad acompaña a este tema fascinante y difícil de abordar, al igual que ocurre con la conciencia.

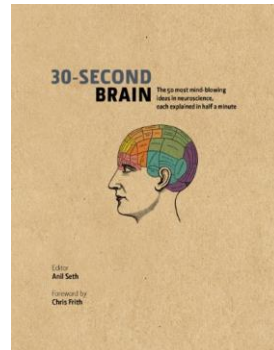
Libros de Anil Seth



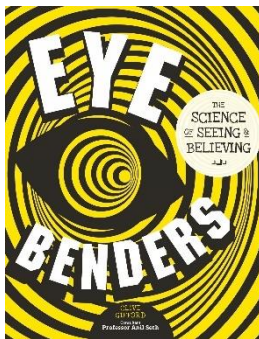
2021



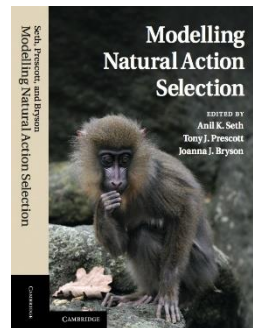
2015



2014



2013



2011

VII. Diálogo con Karl Friston



Karl J. Friston es profesor en el Instituto de Neurología en University College London, siendo un neurocientífico y una autoridad en imagen cerebral, habiendo inventado el mapeo paramétrico estadístico, la morfometría basada en voxels y la "Modelización Causal Dinámica" (DCM).

Fue galardonado con el "Minerva Golden Brain Award" y elegido como "Fellow de la Royal Society". Recibió la "Medalla del Collège de France" en 2008 y el premio "Weldon Memorial" en 2013 por sus contribuciones a la biología matemática.

En 2016, recibió el premio "Charles Branch" por descubrimientos sin igual en investigación cerebral y el premio "Glass Brain". Friston cuenta con nombramientos honoríficos de la Universidad de Zurich y la Universidad Radboud.

Publicó varios libros, como: *Statistical Parametric Mapping: The Analysis of Functional Brain Images* (2006), *Principles of Brain Dynamics: Global State Interactions* (2012) y más recientemente *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior* (2022).

Más información: <https://www.fil.ion.ucl.ac.uk/~karl/>

Pregunta: El Profesor Friston es considerado la máxima autoridad mundial en el llamado “Principio de la Energía Libre” (PEL) y en sus múltiples consecuencias teóricas y empíricas. Mi primera pregunta, de hecho, es una meta pregunta: ¿Cómo surgió esta idea del “Principio de la Energía Libre” en su investigación?

Karl Friston: Podemos argumentar que la idea del “Principio de la Energía Libre” (*Free Energy Principle*) existe desde los tiempos de Platón, también está presente en el pensamiento de Kant, y fue muy bien articulada por Helmholtz con la noción de inferencias inconscientes.

También podemos encontrarlo en las ideas de Richard Gregory, en el supuesto de la percepción como prueba de hipótesis y formalizaciones, así como en personas como Peter Dayan y Jeffrey Hinton y su “Máquina Helmholtz”. Estos autores fueron inspirados por la noción matemática de energía libre de Richard Feynman, quien también intentaba usar el principio para convertir un problema de integración de inferencia insoluble en un problema de optimización.

Esto significa que podemos considerar el cerebro como una máquina de optimización, que constantemente intenta optimizar sus creencias. Entonces, podemos interpretar el cerebro como una máquina de inferencia, un órgano estadístico.

Estas ideas han sido desarrolladas durante literalmente milenios, pero fueron formalizadas por el trabajo de personas como Richard Feynman, y en términos de Inteligencia Artificial y aprendizaje automático, por personas como Jeffrey Hinton y Peter Dayan.

Las generalizaciones de estas ideas implican varias etapas. La primera generalización sería aplicar exactamente las mismas ideas no solo para hacer inferencias, mediciones y observaciones, sino también considerar que nuestro comportamiento y nuestras acciones pueden entenderse como un intento de minimizar la energía libre.

Si usamos, por ejemplo, el aprendizaje automático como una especie de energía libre, que es una energía libre negativa, si estuviera hablando con físicos, estarían hablando de minimizar la energía libre o minimizar el error de predicción, minimizar la sorpresa.

Siguiendo esto, ahora podemos pensar en todo lo que hacemos como intentos de minimizar la sorpresa, en el sentido de que anticipamos nuestros movimientos utilizando, por ejemplo, una teoría ideomotora para hacer estas predicciones. Todo esto puede entenderse como minimizar exactamente la misma energía libre, error de predicción o sorpresa.

Por lo tanto, con estas y otras influencias, fui formalizando gradualmente estas ideas, cristalizándolas a lo largo de un largo período de

tiempo hasta la actualidad, tratando de desarrollar un enfoque relevante para explicar el funcionamiento del cerebro humano.

Pregunta: Según el “Principio de la Energía Libre”, los organismos vivos buscan minimizar la entropía de su energía libre. Siguiendo esta premisa, ¿Cree que podría haber un problema al considerar, por ejemplo, estados alterados de conciencia, ya que generalmente implican una entropía anormalmente alta o baja? ¿Considera que estos casos podrían ser problemáticos desde su perspectiva? Estoy pensando, por ejemplo, en estados inducidos por drogas u otros tipos similares de estados inducidos.

Karl Friston: No creo que eso sea problemático. Creo que son ventanas muy importantes y muy informativas para esta visión del cerebro mientras intenta dar sentido a este mundo y avanzar hacia la sensación activa.

Creo que tanto la psicopatología como los diferentes estados mentales que podemos experimentar al consumir ciertas drogas psicodélicas revelan aspectos mecanicistas muy importantes desde esta perspectiva.

Steven mencionó la entropía. La entropía es simplemente una descripción de una distribución de probabilidad, por lo tanto, es muy importante

determinar a qué distribución de probabilidad se refiere en este caso.

El “Principio de la Energía Libre”, cuando se aplica a cualquier tipo de autoorganización, desde pequeñas partículas hasta seres humanos, la entropía a la que las personas se refieren y que se está enfocando es la entropía de los resultados. Son las cosas que realmente experimento como organismo.

Esta sería la perspectiva de la homeostasis, el organismo está intentando mantener su interacción con el entorno dentro de límites viables.

Entonces, el organismo está intentando minimizar esa entropía, pero al mismo tiempo, para lograrlo, también debe maximizar la entropía de sus creencias de acuerdo con principios como el de Occam. Queremos encontrar explicaciones simples sin comprometernos demasiado con una explicación particular.

Podemos observar que hay una dialéctica entre intentar hacer el mejor tipo de inferencia, lo cual implicaría maximizar la entropía, y la consecuencia de ello, que es minimizar la entropía física de mis resultados.

El mismo argumento también se aplica a la entropía de nuestra actividad neuronal que codifica nuestras creencias sobre las causas de nuestras sensaciones. Entonces, al mismo tiempo, el cerebro está maximizando la entropía de las creencias, pero también

está minimizando en gran medida la entropía de la dinámica neuronal, de modo que sea mínimamente compleja, intentando así hacerla lo más eficiente posible.

Creo que debemos tener cierto cuidado al hablar de la entropía, porque existen diferentes tipos de entropías, como la entropía de la actividad general, la entropía de nuestras creencias y la entropía de las observaciones. Pero todos estos diferentes tipos de entropía son formas muy útiles, como destacó Steven, para resumir diferentes estados de conciencia.

La historia en relación con condiciones psiquiátricas y/o drogas que alteran la mente, y de hecho, diferentes estados alterados de conciencia que son fisiológicos, como estados de sueño, por ejemplo, o incluso sueños lúcidos, todos estos estados se centran en aspectos muy interesantes de este proceso de optimización de la energía libre.

A veces, con su codificación predictiva, si el organismo hace algunas suposiciones simplificadoras, se puede interpretar la energía libre como un error de predicción, más precisamente, un error de predicción ponderado con precisión. En este caso, no se intenta minimizar todos los errores de predicción, sino solo aquellos que se consideran muy precisos, informativos y confiables.

Cuando pensamos en el cerebro como si estuviera siendo impulsado por el imperativo de minimizar

errores precisos de predicción ponderada, nos damos cuenta de que no es suficiente hacer predicciones sobre el contenido de las observaciones, sino también sobre su precisión.

Es necesario también cuantificar su incertidumbre y lo que eso significa desde el punto de vista de un fisiólogo es que se debe estimar la excitabilidad de varias estructuras neuronales al reportar errores de predicción a niveles superiores que son responsables de acumular evidencia y actualizar creencias.

Dicho de manera muy simple, lo que esto significa es que la modulación de la excitabilidad de las estructuras neuronales que informan sobre errores de predicción, al informar sobre la energía libre, debe ser anticipada por sí misma. Y si el cerebro comete errores en estas predicciones, obtendrá algunas creencias e inferencias muy extrañas y anormales.

Esto es importante por dos razones centrales: en primer lugar, se conecta con los aspectos psicológicos de la ponderación de la decisión, que sería la atención. Prestar atención a algo básicamente implica privilegiar aquellos que predicen errores, teniendo acceso preferencial a la actualización de ciertas creencias pero no de otras, algo así como: "voy a ignorar esto, no es novedoso y tiene baja precisión; son noticias falsas. Pero esto parece ser interesante, estoy evaluando que esta información sea importante o precisa; voy a

aumentar la ganancia de atención hacia estos errores de predicción y actualizar mi perspectiva."

Desde el punto de vista farmacológico, los medicamentos que tienen estos efectos son exactamente los mismos que afectan la excitabilidad de esta ponderación de precisión. Por lo tanto, ahora podemos establecer una conexión interesante entre el papel de los psicodélicos como la psilocibina y/o el LSD, y los informes sobre la atención visual que explican por qué obtenemos inferencias perceptivas anormales bajo el efecto de estos psicodélicos.

Pero lo interesante es que no necesitamos tomar psicodélicos para tener esas inferencias anormales. Podemos observar exactamente las mismas fallas en la neuromodulación como fenotipos de ciertas condiciones psiquiátricas que conducen a inferencias falsas.

Con esta "inferencia", quiero decir simplemente los errores clásicos de "tipo uno" y "tipo dos": inferir algo que está presente cuando no lo está, como una alucinación, o inferir que algo no está presente cuando sí lo está, lo que sería una síndrome disociativa o síndrome de negligencia. Ambos son características de muchas condiciones neurológicas y psiquiátricas que pueden explicarse mediante mi enfoque.

Pregunta: Mi próxima pregunta es precisamente cómo podemos aplicar este enfoque a los trastornos mentales, como la esquizofrenia o el autismo. Creo que la fuerza de sus ideas radica en el hecho de que podemos explicar tanto el cerebro "normal" o saludable, como también el cerebro anormal, los estados alterados de conciencia e incluso los sueños lúcidos. ¿Considera que esta abundancia explicativa es una señal de que va por el camino correcto?

Karl Friston: Sí, el caso específico de los sueños lúcidos es muy interesante.

Pregunta: Creo que el profesor Friston tiene un artículo publicado donde aborda los sueños lúcidos desde su perspectiva de inferencia activa y energía libre, del año 2018⁸¹ si no me equivoco.

Karl Friston: Sobre los sueños, puedo mencionar brevemente a mi colega y amigo Allan Hobson, con quien estudié hace algunos años y que estaba fascinado por el fenómeno de los sueños lúcidos. También se centraba en destacar que, si buscamos evidencia de que el cerebro es un órgano constructivo en el sentido de que genera fantasías o hipótesis sobre cómo funciona el mundo, entonces no necesitamos ir mucho más allá que enfocarnos en los sueños.

⁸¹ Publicación original: Hobson, A. & Friston, K. (2014) "Consciousness, Dreams, and Inference", *Journal of Consciousness Studies*, 21 (1-2): 6-32.

Durante el sueño, experimentamos esta desconexión neuroquímica/farmacológica/neuromoduladora de la precisión de todas nuestras sensaciones, excepto las de los ojos, ya que estos forman parte del sistema nervioso central.

Y aún así, seguimos teniendo comprensión, seguimos soñando. Así que creo que este es un ejemplo maravilloso del cerebro como un órgano activo, creativo y constructivo, que se contextualiza únicamente al obtener esa ponderación precisa y ese tipo de control de ganancia en el equilibrio adecuado.

Pregunta: Tengo una provocación filosófica para el profesor. En algunos artículos, parece defender una visión corporizada o anti-representacionista. Sin embargo, en otros artículos parece argumentar a favor de una visión representacionista. ¿Cuál es su posición real sobre lo que se conoce como la "guerra de la representación" en la Ciencia Cognitiva y Filosofía de la Mente? ¿El profesor está a favor o en contra de las representaciones en general?

Karl Friston: Mi respuesta es la siguiente: Depende un poco de quién sea mi interlocutor. El otro día escuché una frase bonita en una comedia australiana sobre querer "agradar" a las personas.

Si estoy hablando con un escéptico, alguien que es anti-realista, el "Principio de la Energía Libre" puede ser

utilizado para defender esa posición: "Sí, es absolutamente cierto". ¿Por qué? Porque el organismo nunca estará directamente expuesto a lo que está "afuera", asumiendo que haya algo fuera.

Todo a lo que tenemos acceso es el sensorium de la experiencia. Somos un cerebro "encerrado" en una caja y todo lo que creemos sobre el mundo es solo una fantasía creada por él. Entonces, este argumento sería muy consistente con el enfoque escéptico de la representación.

Por otro lado, si estoy hablando con un realista o un externalista, la interpretación de la inferencia activa o la aplicación del "Principio de la Energía Libre" celebrará esto, porque todos estos aspectos están relacionados con la sensación activa corporal, se trata de cómo "yo" me involucro físicamente con mi mundo para obtener el tipo adecuado de sensaciones que hagan que mis creencias internas sean más consistentes con ese mundo.

Por lo tanto, podemos interpretar la auto-evidencia literalmente como la recopilación de evidencia del mundo real para el modelo-"yo" de ese mundo. Si "yo" soy ese modelo, estoy recolectando pruebas de mi existencia y esta es una forma de considerar los imperativos existenciales implicados por este tipo de auto-evidencia.

Entonces, el “Principio de la Energía Libre” y la “inferencia activa” no admitirían un enactivismo radical. Sin embargo, si se trata de un enactivismo más suave, creo que la teoría puede acomodarlos fácilmente.

Pregunta: ¿Qué consecuencias tiene el “Principio de la Energía Libre” en relación con lo que realmente buscan los filósofos, es decir, en relación con la existencia de las cualidades subjetivas de la experiencia, los llamados “qualia”? ¿Su enfoque proporciona algún tipo de conocimiento relevante para abordar esta noción filosófica?

Karl Friston: Esta es una pregunta muy interesante y desafiante, que ha sido un enfoque reciente entre varios de mis colegas que se dedican a la fenomenología de la subjetividad.

Mi mejor resumen del pensamiento actual relacionado con las personas que conozco que se centran en la filosofía de la mente y tratan de enmarcar la subjetividad dentro del “Principio de la Energía Libre” y la “inferencia activa” sería que, para tener una experiencia cualitativa, es necesario que sea una acción interna (acción encubierta).

Si yo fuera un psicólogo, sería equivalente a decir que debo poder prestar atención a algo específico. Si no puedo elegir en qué enfocar mi atención, entonces no puedo tener la experiencia cualitativa de eso.

Muchas veces, esto está vinculado a las nociones de transparencia y opacidad fenomenales que el filósofo Thomas Metzinger aborda, por ejemplo. Entonces, para que algo se vuelva opaco —y ahora estoy "leyendo" esa opacidad como una especie de isomorfo con una calidad de experiencia de una percepción— es necesario tener la capacidad de alterar la precisión de la evidencia sensorial que es relevante, es decir, el error de predicción.

Pero si fuera un neuroanatomista, diría que tener experiencias cualitativas —o al menos tener esos tipos de percepciones que pueden volverse opacas— significa que debería poder encontrar proyecciones neuronales anatómicas en el cerebro que tengan un efecto neuromodulador en algunas partes más internas del mismo en una jerarquía determinada.

Y, si volviera a ser psicólogo, esperararía que viéramos esas proyecciones respaldadas por la capacidad de atender, de forma endógena, a diferentes partes de lo sensorial. Así que lo que tenemos aquí es una historia, una historia que conecta una especie de calidad de experiencia con la atención, con la selección activa de fuentes de evidencia de mis fantasías y la actualización de mis creencias. Para mí, la selección activa no es un tipo de comportamiento obvio.

Podemos ver paralelos, sin duda, con la teoría premotora de la atención de Rizzolatti, en el sentido de que actuar en el mundo para obtener el tipo adecuado

de información precisa puede, de hecho, implicar una acción abierta mientras exploro visualmente el mundo.

Pero si internalizo esto y ahora pienso en esa exploración de lo sensorial internalizándose a través de varias fuentes selectivas de errores de predicción o evidencias sensoriales, fechándolas con el control de precisión, eso solo puede ser un tipo de acción mental. Para mí, esto es una prueba decisiva para la calidad de la experiencia, o la interpretación de los qualia como experiencia cualitativa.

Si estamos de acuerdo con esta interpretación, hay algunas condiciones que deben cumplirse antes de que podamos participar en este tipo de acción mental: es necesario tener un modelo generativo que pueda generar predicciones de precisión que se utilicen como base para la acción mental, lo que nos indica de inmediato que se requiere tener un modelo generativo del futuro.

Esto dice mucho sobre el tipo de criaturas, sistemas o partículas que podrían posiblemente tener experiencias cualitativas, tienen que ser aquellas criaturas bastante sofisticadas que poseen modelos generativos del futuro, porque es necesario tener un modelo de las consecuencias de la acción, incluso si esa acción es secreta o interna (encubierta).

Creo que esto es muy importante. Hay una dimensión temporal implicada en este tipo de explicación

mecanicista de la experiencia cualitativa. Otros argumentos que colegas, especialmente personas como Maxwell Ramstead, Mark Solms y otros, han escrito recientemente defienden una formulación más abstracta de la conciencia.

Estos colegas intentan pensar en la experiencia cualitativa como proyectada en una "Manta de Markov" desde el punto de vista de la teoría clásica de la Información, o como un holograma desde el punto de vista de la "Teoría Cuántica de la Información".

La idea básica es que quizás un sistema físico distribuido en el cerebro pueda ser subdividido aún más a través de la "Manta de Markov" que envuelve ese sistema —o, si se adopta la visión de un teórico de la información cuántica, el holograma que contiene la información clásica que los aspectos internos de esta estructura producen— pueda separar esa estructura interna del resto del cerebro.

Hay argumentos que sostienen que, para que algo pueda ser consciente y tener experiencias conscientes, se requiere la existencia de esta "Manta de Markov" o de una única pantalla holográfica, lo cual, curiosamente, nos remite de nuevo a un "Teatro Cartesiano"

Sin embargo, no hay un dualismo esencial implícito en esta visión, pero nos lleva de manera interesante de vuelta a la noción de una pantalla interna. Lo crucial en

esta interpretación es que los estados internos que observan esta pantalla nunca podrán verse a sí mismos, porque no es posible insertar otros flujos internos.

Esta historia puede ser revelada en el contexto de la atención y la acción interna, o como una historia de acción mental, simplemente porque la única forma en que el interior puede cambiar las proyecciones en esta pantalla interna es a través de esta acción mental. Pero siempre debemos regresar a esa noción de acción mental al final del día, no podemos escapar de ella.

Pregunta: Para concluir, no estoy seguro si el profesor Friston está al tanto del nuevo libro "The Model of the Mind" de Grace Lindsay, una neurocientífica computacional de la Universidad de Nueva York, que argumenta que su enfoque basado en el "Principio de la Energía Libre" es infalsificable desde el punto de vista de la filosofía de la ciencia, es decir, que no es posible determinar científicamente si el principio es verdadero o no. ¿Qué opina de esta crítica?

Karl Friston: Creo que esta crítica puede tener todo el sentido: el Principio de la Energía Libre es, de hecho, infalsificable. ¿Por qué? Es muy simple. Este principio es un principio de "mínima acción", como el principio de "mínima acción" de Hamilton. En física, matemáticas e incluso en ciencias de la computación, este tipo de principios son solo métodos o herramientas, no son

teorías ni hipótesis; son simplemente herramientas útiles.

El “Principio de la Energía Libre” es simplemente eso, una herramienta, no está ni correcto ni incorrecto; o el investigador aplica este principio, o no lo aplica. Su propósito es ser aplicado, simulado o reproducido, como una creación de sentido o comportamiento sensible.

A veces implica combinar simulaciones u observar el comportamiento para poder reproducir realmente el comportamiento observado. En esencia, es un principio o método que permite simular o reproducir un comportamiento básicamente consciente.

Pero al hacerlo, es necesario comprometerse a aplicarlo a un modelo generativo específico. En ese punto, si se afirma que cierto modelo generativo es capaz de explicar este tipo de comportamiento, este tipo de criatura o este paciente en particular, entonces eso se convierte en una hipótesis y es difícil de falsificar.

Lo que se hace en este caso es aplicar el “Principio de la Energía Libre” a varios modelos generativos de una determinada hipótesis para descubrir cuál es, efectivamente, el mejor modelo para esa hipótesis.

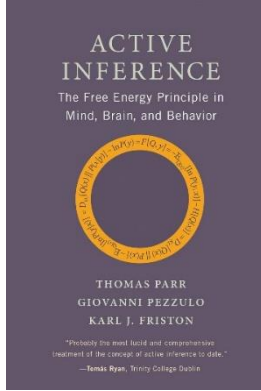
Pero si se interpreta el modelo generativo aplicando el “Principio de la Energía Libre” como una teoría, entonces ciertamente se vuelve falsificable, podría existir un modelo mejor que explique el

comportamiento de una persona determinada o un tipo específico de comportamiento. Sin embargo, el principio en sí mismo es solo eso, un principio, no puede ser falsificado por sí solo.

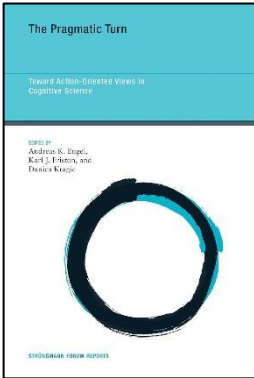
Para concluir, lo que el “Principio de la Energía Libre” ofrece es una estructura teórica formal en la cual podemos entender la noción de falsificación. Se trata simplemente de comparar las evidencias de la hipótesis nula con respecto a la hipótesis alternativa.

Por eso no hay ningún problema en afirmar que, desde un punto de vista particular, el “Principio de la Energía Libre” es infalsificable, y es por eso que digo que la autora tiene absolutamente razón.

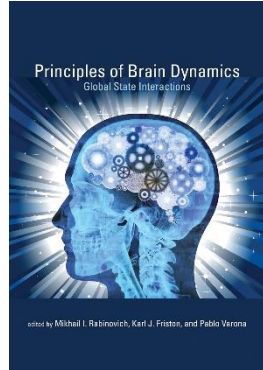
Libros de Karl Friston



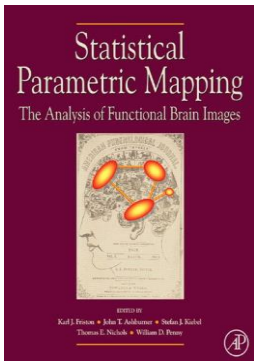
2022



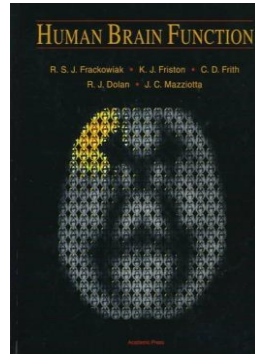
2016



2012



2011



1997

VIII. Diálogo con Christof Koch



Christof Koch es un neurofisiólogo y neurocientífico computacional conocido principalmente por su trabajo sobre las bases neurales de la conciencia.

Es el presidente y jefe científico del “Allen Institute for Brain Science en Seattle” y es el jefe científico de la fundación “The Tiny Blue Dot”.

De 1986 a 2013, fue profesor en el “Instituto de Tecnología de California”. El principal colaborador de Koch en el esfuerzo por localizar los correlatos neuronales de la conciencia fue Francis Crick, Premio Nobel de Medicina.

Más recientemente, Koch ha trabajado en estrecha colaboración con el psiquiatra y neurocientífico Giulio Tononi y ha estado desarrollando la “Teoría de la Información Integrada” de la conciencia.

Koch es autor de varios libros, tales como: *The Quest for Consciousness: a Neurobiological Approach* (2004), *Consciousness: Confessions of a Romantic Reductionist* (2012) and *The Feeling of Life Itself - Why Consciousness is Widespread but Can't be Computed* (2019).

Más información: <https://christofkoch.com/>

Pregunta: Sabemos que la conciencia ha planteado grandes dificultades a la ciencia contemporánea. ¿El profesor Koch cree que necesitamos una especie de redefinición de la naturaleza física fundamental para explicar la conciencia? ¿O la física actual y el fisicalismo son suficientes para abordar la conciencia?

Christof Koch: Depende de lo que entendamos por fisicalismo. Apoyo una reformulación de la "Teoría de la Información Integrada" (TII) que sostiene que no es solo el poder causal extrínseco descrito por la física convencional, sino que también es necesario investigar el poder causal intrínseco, porque eso es lo que es la conciencia. Esto no significa que necesitemos una nueva teoría integrada de la mecánica cuántica y la gravedad, como defiende Sir Roger Penrose.

Tendremos que ver hacia dónde evolucionará la Física, pero en este momento, dado que el cerebro funciona a 300 grados Kelvin, no creo que, incluso con una reformulación que logre una teoría única y completa de las leyes físicas, esto haga una diferencia para el cerebro, ya que los cerebros no operan en esa escala. Puedo estar equivocado, pero esa es mi intuición.

Pregunta: La "Teoría de la Información Integrada" (TII) está compuesta por cinco axiomas, pero algunos autores critican la inclusión del "Principio de la Exclusión", considerándolo un tipo de principio

arbitrario o un mecanismo ad hoc. El profesor Koch opina que este principio de exclusión es un punto débil de la TII que necesita ser mejorado, o cree que el hecho de que cualquier experiencia consciente se defina por sí sola es suficiente para respaldar la inclusión del principio de exclusión?

Christof Koch: Discrepo totalmente de esa crítica porque creo que ese principio es esencial. ¿Por qué? Porque de lo contrario, obtendríamos una multiplicidad de experiencias de conciencia y, como consecuencia, una multiplicidad de poderes causales. Por lo tanto, es una parte esencial de la teoría.

Este principio de exclusión no fue añadido posteriormente ni de manera ad hoc, sino que siempre ha formado parte de la teoría porque, en última instancia, dado toda la combinatoria posible, hay un número muy grande de mecanismos posibles que existen por sí solos. La TII afirma que solo hay uno posible, el que representa el máximo de toda la granularidad espacial y temporal.

Esto significa que solo hay una conciencia y no puede haber un número infinito de conciencias en mi cabeza. ¿Por qué esta escala espacial y temporal específica está teniendo esta relación particular en mi cerebro?

La respuesta a esa pregunta es la siguiente: ¿Por qué es esa relación particular la que maximiza el poder causal intrínseco?. Si tuviera neuronas adicionales, o si tuviera

menos neuronas, tendría menos poder causal. Si observamos en una escala de tiempo diferente, tendría menos poder causal. Esa es la afirmación de la teoría.

La teoría podría estar equivocada, pero no se puede eliminar arbitrariamente un axioma y esperar que toda la teoría siga siendo válida.

Pregunta: Hubo un pequeño cambio o actualización de la versión inicial de la TII que ahora llamamos la versión 3.0 de la TII, donde hay un intento de redefinir la conciencia como el máximo poder causal irreducible de cualquier red. El profesor cree que sería más acertado pensar en la teoría como centrada en poderes causales integrados en lugar de información integrada en sí misma.

Christof Koch: Quizás, eso me parece un argumento lingüístico. En última instancia, se trata de conciencia. Bueno, si pensamos un poco más, podríamos decir que son ambas cosas. La afirmación es que el poder causal intrínseco es lo que constituye la conciencia. Como mencioné, es una identidad.

En la TII, etiquetamos esta información integrada como Φ , pero la teoría es más que solo Φ , porque cualquier teoría de la conciencia no solo debe explicar cómo la conciencia encaja en el orden natural de las cosas, sino también por qué la conciencia del dolor es diferente de la conciencia del placer o de la percepción del espacio.

Una buena teoría de la conciencia debe poder explicar la fenomenología y por qué ciertos estados nos hacen sentir lo que sentimos, y creo que esta teoría de la TII logra precisamente eso.

Pregunta: Recientemente, hubo una controversia en el estudio científico de la conciencia donde varios científicos y filósofos firmaron una carta pública argumentando que la TII era una teoría pseudocientífica, no falsificable. ¿Qué tipo de experimento cree usted que podría falsificar su teoría de la TII?

Christof Koch: Muchas de ellas, de hecho, la TII realiza pruebas y predicciones concretas. Por ejemplo, podemos observar una red neuronal específica y preguntarnos si maximiza el poder causal intrínseco. Es una predicción muy específica, después de tener la matriz de probabilidad de transición de esa red neuronal, en principio, podemos calcularla. Y luego podemos preguntarnos: ¿Esta red maximiza el poder causal? ¿La fenomenología de este sistema realmente lo explica?

Me refiero brevemente a este tema en un artículo⁸² sobre la fenomenología del espacio, el flujo del tiempo

⁸² Artículo publicado aquí: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0268577>.

y el color. Actualmente, hay un consorcio⁸³ que busca colaborar para probar predicciones entre la TII y la “Teoría del Espacio de Trabajo Global” (TETG). La TETG sostiene que el sustrato de la conciencia se encuentra en el parietal prefrontal, mientras que la TII indica que está en la parte posterior del cerebro.

Por lo tanto, en este momento estamos probando científicamente varias predicciones de la teoría. La TII afirma que un organismo estará consciente mientras esté experimentando: si experimenta algo durante 10 segundos, habrá un sustrato físico de esa experiencia consciente durante 10 segundos. La TETG sostiene lo contrario, el sustrato solo está presente cuando envía la transmisión por primera vez y luego desaparece. Nuevamente, hay una amplia variedad de formas en las que esta teoría puede ser probada.

Otro ejemplo, existe un método de prueba de complejidad de perturbación que permite evaluar, en un contexto clínico, cuándo el paciente está consciente o no. Esto proviene directamente de la TII y esta predicción está siendo probada actualmente en ensayos clínicos con seres humanos.

⁸³ Por ejemplo: Melloni L, et al. (2023) “An adversarial collaboration protocol for testing contrasting predictions of global neuronal workspace and integrated information theory”, *PLOS ONE*, 18 (2): e0268577.

Pregunta: Eso es una buena señal, ¿Verdad? Porque muchas teorías de la conciencia no pueden ser probadas, por lo tanto, esto parece ser una ventaja de la TII en comparación con otras teorías de la conciencia.

Christof Koch: Correcto. Diría que es incluso esencial.

Pregunta: En relación con la "Inteligencia Artificial", el Profesor Koch argumenta que los ordenadores, al menos aquellos que utilizan una arquitectura de Von Neumann, nunca podrán ser conscientes. Sin embargo, ¿Estaría de acuerdo en que si los poderes causales intrínsecos del cerebro fueran reproducidos en un sustrato artificial de silicio, por ejemplo, podríamos asumir entonces que serían conscientes?

Christof Koch: Me parece que eso es plausible. No hay nada mágico en el cerebro, no hay nada sobrenatural en el cerebro. Un transistor típico puede estar conectado con dos o tres receptores. En cambio, un neurona está conectada con 50,000 otras neuronas.

Además, hay una gran superposición entre estos 50,000 neuronas. Por lo tanto, considerando dos neuronas cercanas, una proyecta hacia 50,000 neuronas, y la otra también proyecta hacia 50,000 neuronas, pero se superponen en gran medida entre sí.

Novamente, todo esto es muy diferente de la conectividad, que en última instancia determina su

poder causal. Entonces, si incorporamos esa fuerza causal en hardware, sea cual sea la materia de ese hardware, podríamos obtener conciencia, sí.

Mas debemos tener mucho cuidado aquí: el concepto de "inteligencia" es diferente del de "consciencia". De hecho, son dos cosas muy diferentes. No hay duda de que alcanzaremos inteligencia "artificial", es decir, una "Inteligencia Artificial General" (AGI), ya que en última instancia se trata simplemente de hacer algo. Pero la consciencia no es algo que se hace, sino algo que es: la consciencia es un estado de ser. Por otro lado, la inteligencia, en última instancia, se trata de hacer. No hay problema en poner inteligencia humana o sobrehumana en máquinas, pero eso no significa necesariamente que sean conscientes.

Pregunta: El profesor Koch mencionó una relación entre la Inteligencia Artificial y la consciencia, afirmando que la "Inteligencia Artificial" es inevitable, pero no estamos seguros de si será consciente o no. Mi pregunta es: ¿Es realmente posible que la inteligencia exista sin consciencia? ¿La inteligencia no viene después de ser consciente? ¿Cómo puede la Inteligencia Artificial existir sin ser consciente?

Christof Koch: Esa es una buena pregunta. El único ejemplo que tenemos en este momento de verdadera

inteligencia flexible es la inteligencia humana, y de hecho somos conscientes.

No obstante, muchos de nosotros creemos que esta es solo una forma de volvernos inteligentes, una forma que la evolución "eligió", pero que existen otras formas, quizás a través de silicio y software. Hasta ahora, no veo ninguna evidencia de que esas otras formas puedan funcionar.

Vamos a plantear la cuestión de otra forma. Puedo imaginar fácilmente cuán más poderosos pueden ser los ordenadores, o las máquinas de Turing, que yo, ciertamente mucho más rápidos que yo, que son inteligentes o incluso superinteligentes. No veo, sin embargo, a priori, por qué la consciencia sería necesaria para eso.

Si pensamos en "aprendizaje profundo" (deep learning), en redes generativas o en redes de transformadores, etc., esas tecnologías parecen estar desarrollándose muy bien; cuanto mayor es la escala, mejor es su desempeño. No veo por qué la consciencia sería relevante para estas tecnologías.

Quién sabe si, entretanto, esas tecnologías encontrarán algunos obstáculos. Pero mi suposición es que podemos llegar a la "Inteligencia Artificial General" (AGI), incluyendo la superinteligencia, sin involucrar necesariamente la consciencia.

La evolución no produjo inteligencia o consciencia de esta forma, pero ahora, como seres humanos, podemos hacer las cosas de manera diferente a como lo hizo la evolución. Si eso es bueno para nosotros, es una cuestión totalmente diferente.

¿Es realmente bueno para el Homo Sapiens-Sapiens que existan estas superinteligencias? Soy muy escéptico al respecto, pero es el viaje que estamos haciendo en el presente, para bien o para mal.

Pregunta: Existen varios métodos para estudiar el cerebro consciente. Podemos centrarnos en los trastornos de la consciencia, en metodologías basadas en psicodélicos, o también en la estimulación cerebral, etc. ¿Cuál de estos métodos cree que es el más prometedor para estudiar la naturaleza de la consciencia?

Christof Koch: Diría que todos esos métodos son relevantes, no hay ninguno que sea específicamente mejor, ya que depende del historial y la formación del investigador, de su experiencia, etc.

Todos los diferentes métodos tienen ventajas y desventajas. Algunos son más objetivos, algunos pueden incorporar datos de primera persona, mientras que otros son más difíciles de incluir esos datos.

De hecho, podemos hacer preguntas relevantes a las personas cuando intentamos estudiarlas, pero por otro lado, no podemos intervenir en los cerebros de las personas como lo hacemos con los cerebros de los animales en la experimentación científica.

Por lo tanto, cada una de estas técnicas tiene sus ventajas y desventajas. Realmente depende del interés individual, de su formación y de qué ángulo desea estudiar la consciencia.

Los psicodélicos, por ejemplo, tienen la ventaja de poder impactar enormemente en la consciencia, pero también tienen desventajas. La mayoría son ilegales y, lo que es más importante, no sabemos mucho sobre su naturaleza.

Muchos estudios se han centrado en los psicodélicos en ratas, pero es muy difícil saber realmente qué están experimentando subjetivamente cuando consumen algún psicodélico específico. Por lo tanto, diría que depende de su formación y de sus intereses profesionales.

Pregunta: Pensando en las teorías del procesamiento predictivo, ¿Qué piensa usted sobre las influencias "de arriba hacia abajo" (top-down) en el cerebro? ¿Cree que son relevantes para entender la naturaleza de la consciencia?

Christof Koch: En la mayoría de las personas, dejando de lado a los esquizofrénicos, lo que cada persona experimenta es bastante estable. Podemos ser confrontados con ilusiones visuales durante algunos minutos y seguir viendo esa misma ilusión.

En mi opinión, esto me dice que las influencias "de arriba hacia abajo" de mis expectativas no son tan fuertes. Definitivamente existen, no hay duda, especialmente cuando tengo poco tiempo para procesar información, como cuando tengo 100 milisegundos para reportar lo que veo en esas condiciones; en esos casos, esas influencias "de arriba hacia abajo" son más importantes.

Pero en general, creo que se enfatiza demasiado ese papel. Puedo mirar una pared y ver la pared. No hay nada que prever, nada cambia, simplemente veo una pared vacía.

Entonces, esto me indica que el enfoque "de arriba hacia abajo" no es tan importante, al menos en estas condiciones, como la gente cree que es. Y no creo que tenga mucho que ver con la codificación predictiva en la mayoría de los casos.

Pregunta: ¿Qué tipo de consecuencias éticas cree que podría tener la TII? Específicamente, en relación con su postura hacia el pampsiquismo, que sostiene que todas las partículas en el mundo poseen algún grado de

conciencia. Tomemos, por ejemplo, la ética animal en general. ¿Podemos asumir que, al menos filosóficamente, una de las predicciones de la teoría es que muchos animales son capaces de sentir dolor y placer, especialmente criaturas con cerebros desarrollados? ¿Cree que esta teoría de la consciencia tiene implicaciones éticas directas?

Christof Koch: Sí, absolutamente. Hace muchos años, precisamente por eso, me convertí en vegetariano. De hecho, ni siquiera mato abejas y trato de no matar más insectos porque también sienten. Realmente creo que muchos de los animales que pensamos que no tienen la capacidad de sentir dolor o placer realmente sienten algo. Pero ellos no tienen voz.

Una abeja, por ejemplo, no tiene voz ni una mente como la mía. Claramente, su cerebro es mucho más simple que el mío, pero también experimenta felicidad cuando simplemente está bebiendo un poco de néctar dorado y volando bajo el sol caliente.

Y al igual que yo, también está atrapada entre dos eternidades, el inicio y el final de su vida. Ambos estamos arrojados a este universo y debemos minimizar el sufrimiento de todas las criaturas, incluyendo criaturas como las abejas y similares.

Por lo tanto, creo que la TII tiene implicaciones éticas claras y debemos actuar de manera coherente con esas

consecuencias, incluso si eso requiere un cambio radical en nuestro estilo de vida.

Pregunta: Con todo el conocimiento que el profesor ha acumulado en las ciencias del cerebro, ¿Podríamos decir que está en una posición ventajosa para corregir, por ejemplo, un mal hábito que tenga en su día a día? ¿Este conocimiento increíble y complejo le brinda acceso a una puerta de entrada para cualquier mal hábito que pueda querer corregir, o más bien este conocimiento no le proporciona ninguna ventaja específica para su vida diaria?

Christof Koch: No, absolutamente no me brinda ninguna ventaja. Tengo que luchar la misma batalla que todas las personas. Estoy intentando beber menos y ser más compasivo con los demás, siguiendo una tradición específica de sabiduría antigua.

Resulta que soy un gran admirador del estoicismo y específicamente de Marco Aurelio, lo cual requiere disciplina y esfuerzo mental constante. La TII no otorga a nadie ningún superpoder que pueda decir "¡Ah, ahora entiendo! Así es como funciona el cerebro; por lo tanto, puedo deshacerme mágicamente de todos mis malos hábitos". No, no funciona así.

Por el contrario, requiere mucha disciplina mental y fuerza de voluntad. Se necesita un gran poder causal para evitar activamente las cosas que pueden

perjudicarnos. De hecho, aquí reside la verdadera libertad según la TII.

La teoría también plantea algunas implicaciones sobre el libre albedrío, afirmando que éste existe en el sentido original, en el sentido de que puedo marcar la diferencia para mí mismo, "soy el dueño de mi destino, soy el capitán de mi alma".

Tengo que reflexionar sobre lo que hago, pienso o reflexiono y llegar a la conclusión de que algo es un mal hábito o un buen hábito, y decidir qué quiero reforzar o evitar. Se requiere una disciplina constante, día tras día, y esto, por supuesto, cambiará en última instancia mi cerebro y, por ende, mi vida.

Así que tengo la libertad de determinar mi propio destino, dadas las condiciones de fondo sobre las cuales, evidentemente, puedo influir mucho menos.

Pregunta: Parece estar muy de moda un enfoque "espiritual" de la ciencia, como la sincronicidad cuántica, la curación cuántica, etc., y esto está impactando la vida de cada vez más personas. ¿Cuál es su opinión sobre estos enfoques? ¿Considera que sean inmorales? ¿Hay alguna evidencia sobre la relevancia de este tipo de ideas, o es simplemente algo que vende bien y que las personas sin capacidad para entender caen en estas trampas?

Christof Koch: No hay evidencia en este momento de que la mecánica cuántica microscópica, particularmente el entrelazamiento, que es fundamental para la computación cuántica, juegue algún papel en el cerebro.

Si observamos los computadores cuánticos actuales, la temperatura operativa de un computador cuántico es de 25 o 35 milikelvin, es decir, aproximadamente 50 mil veces más fría que la del cerebro humano.

Por lo tanto, es muy improbable que estos efectos jueguen algún papel en la biofísica o la bioquímica del cerebro. Claro está que muchas personas venden todo tipo de cosas, como curas con cristales, pero esto no nos dice nada; está completamente divorciado de cualquier cosa que el cerebro haga o que constituya algo real.

Pregunta: Creo que el professor Koch tiene un debate interesante con el Dalai Lama sobre espiritualidad y cerebro. ¿Podemos terminar con un breve resumen de lo que concluimos de ese debate?

Christof Koch: Fui dos veces a la India y me encontré y debatí con su santidad, el Dalai Lama. Estuvimos de acuerdo en muchas cosas, como en minimizar el sufrimiento de todas las criaturas conscientes, a diferencia de otros enfoques religiosos como el

catolicismo, que no comparte esa preocupación (yo crecí en la fe católica).

El Budismo, especialmente el Budismo Tibetano, enfatiza que la conciencia es común a todas las criaturas, no solo a los seres humanos que pueden hablar de ella. Sin embargo, divergimos cuando hablamos de la reencarnación.

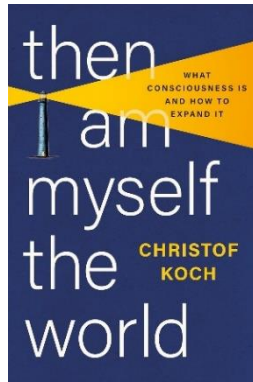
El mantra que propuse consta de cuatro palabras. Sin cerebro, no hay mente ("no brain, never mind"). Esto significa que una vez que el cerebro se disuelve o muere, el sustrato físico de la conciencia desaparece y, en este sentido, ya no hay conciencia sin algún portador físico.

Incluso si se trata de una física exótica, debe haber algún portador del cerebro, sus memorias y sus ideas. Los budistas hablan del "bardo", el espacio liminal entre una vida y su renacimiento en la próxima vida.

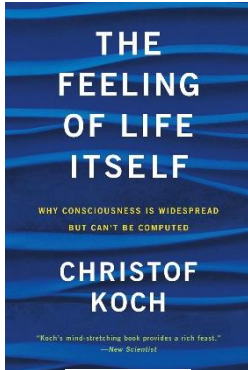
Nuevamente, si algo de mi vida se lleva a la siguiente, debe haber un lugar en el espacio y el tiempo que lleve algún sustrato de mis memorias. De lo contrario, no creo que pueda existir.

Y así terminamos el debate con el Dalai Lama riéndose a carcajadas profundas, diciendo "bueno, lo veremos pronto".

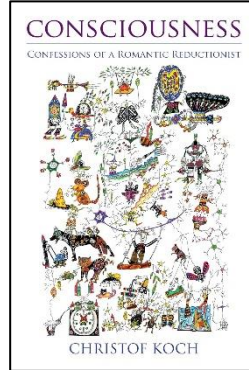
Libros de Christof Koch



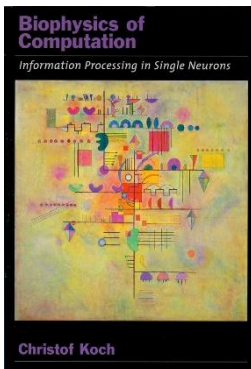
2024



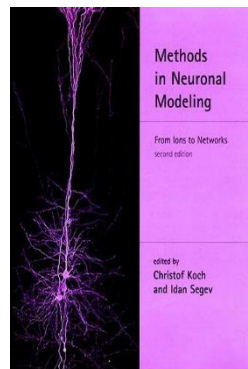
2020



2017



2004



2003

IX. Diálogo con Joseph LeDoux



Joseph LeDoux es profesor de Neurociencia, Psicología, Psiquiatría y Psiquiatría Infantil y Juvenil en la Universidad de Nueva York. Es el Director del "Instituto Cerebral Emocional y del Instituto Nathan Kline de Investigación Psiquiátrica".

Es vice-director del "Centro de Lenguaje, Música y Emoción" en el Max Planck-NYU y miembro de la "Academia Nacional de Ciencias" de Estados Unidos.

LeDoux es autor de varios libros importantes sobre el cerebro, es autor de: *Deep History and Four Realms* (2023), *Anxious* (2015), y *The Emotional Brain* (1996), entre muchos otros. Professor LeDoux es también cantante y compositor para la banda de folk-rock *The Amygdaloids* y el dúo acústico *So We Are*.

Ha ganado varios premios como el "Distinguished Scientific Contributions to Psychology" por la American Psychological Association (2010), el "Karl Spencer Lashley Award" por la American Philosophical Society (2011) y la "Medalla de Gantt" por la Pavlovian Society (2012).

Más información: www.joseph-ledoux.com/.

Pregunta: El profesor LeDoux es uno de los mayores especialistas en el estudio de las emociones y cómo nuestro cerebro, junto con el cuerpo y el entorno, genera ciertos comportamientos. Sin embargo, se muestra escéptico respecto a la utilidad del lenguaje vernáculo o popular (lo que en inglés llamamos "folk language") en relación con las emociones en general. ¿Podría explicarnos más específicamente cómo aborda esta cuestión?

Joseph LeDoux: No es exactamente esa posición la que defiendo, creo que el lenguaje popular (folk) juega un papel importante cuando hablamos de estados mentales, ya que nuestra vida mental se experimenta principalmente en ese lenguaje popular. Sin embargo, establezco un límite cuando aplicamos ese lenguaje específicamente al comportamiento.

Por ejemplo, el cerebro humano puede responder a un estímulo peligroso como un reflejo, o puede tener un patrón de acción fijo más complejo, como la respuesta de congelamiento. También puede, a un nivel más alto, haber aprendido un hábito, o incluso tener un tipo de modelo cognitivo de la situación dirigido hacia un objetivo y responder de esa manera y todo esto puede ocurrir de manera inconsciente.

Pero también puede, en última instancia, responder conscientemente a ese estímulo peligroso. Es decir, mi punto es que no podemos tomar un comportamiento simple como "pelear o huir" (fight or flight) y afirmar

que es, por ejemplo, un indicador de la emoción de miedo, dado que hay muchos comportamientos mediante los cuales el miedo puede expresarse, y no todos ellos implican una experiencia consciente de miedo.

Pregunta: En su investigación, ha estado investigando una parte específica del cerebro llamada amígdala, responsable de detectar y responder a amenazas, pero sostiene que este órgano anatómico no es apropiado para asociarlo directamente con "miedo". ¿Por qué decidió avanzar con un cambio conceptual de lo que comúnmente se llama en neurociencia el "circuito del miedo" hacia lo que el Profesor LeDoux llama "circuito de supervivencia de defensa"?

Joseph LeDoux: Apoyo este cambio conceptual porque realmente no creo que el miedo esté "burbujeando" en la amígdala. Consideremos el circuito de defensa y supervivencia en general, cada animal bilateral, es decir, cada vertebrado que haya vivido, debe tener un circuito que pueda detectar y responder al peligro.

Es un tipo de mecanismo muy primitivo y además de esto, tiene experiencias conscientes asociadas, donde una persona sabe que es ella misma quien está teniendo esa experiencia consciente.

Incluso hice hacer unas camisetas estampadas que dicen "Sin Yo, Sin Miedo" (no self, no fear), lo que

significa que si no se está personalmente involucrado en una forma autoconsciente y autorreferencial, entonces no puede haber emoción.

La persona tiene que formar parte de la experiencia para que algo sea una emoción. Simplemente responder de manera reflexiva o instintiva no explica una experiencia emocional. Si vemos a un perro atropellado, tirado al borde de la carretera, retorciéndose de dolor, todos proyectamos nuestras emociones y sentimientos en ese perro.

Pero lo que realmente estamos observando son "solo" los reflejos del animal cuando se retuerce y gruñe, y así sucesivamente. En mi opinión, no estamos viendo indicadores de dolor efectivo. No estoy, por supuesto, diciendo que el perro no esté sintiendo dolor en ese momento; más bien, estoy afirmando que debemos trazar una línea y separar esas respuestas automáticas y reflejas de las respuestas asociadas a las experiencias conscientes de dolor o miedo, ya que hay una diferencia profunda entre ellas.

Pregunta: Esta visión parece criticar las líneas tradicionales de investigación científica sobre las emociones. Por ejemplo, cuando investigamos con animales, los científicos tienden a observar el comportamiento e intentan etiquetar ese comportamiento con una emoción humana específica.

¿El Profesor LeDoux cree que esta es una forma incorrecta de investigar las emociones en animales?

Joseph LeDoux: Este es un punto interesante. Estaba escribiendo sobre este tema hace algún tiempo y recordaba la referencia bibliográfica que estaba utilizando, un capítulo de un libro sobre antropomorfismo de Elizabeth Knoll, donde aborda la perspectiva de Darwin.

La época en la que Darwin vivió fue la era de la Inglaterra victoriana y el antropomorfismo era parte de esa cultura, era una especie de "modo de vida" de aquel tiempo.

Darwin enfrentaba muchos problemas para que su teoría de la evolución/selección natural fuera aceptada debido a las implicaciones religiosas y tomó la decisión explícita de hablar sobre mentes animales en términos humanos, en lugar de mentes humanas en términos animales, porque pensaba que la segunda opción no sería bien recibida.

Las teorías de Darwin son realmente el punto de partida para el estudio moderno de las emociones, ya que sus seguidores y acólitos a finales del siglo XIX veían el comportamiento como un embajador de la mente.

Aunque esto es ciertamente verdad hasta cierto punto, no es una indicación clara, porque los comportamientos que estudiamos en los animales tienden a ser tipos de respuestas reflejas inactivas y no

respuestas que necesariamente son productos de la experiencia consciente del miedo.

La atribución libre de explicaciones conscientes sin evidencia alguna fue lo que desencadenó toda la revolución conductista en la psicología. No se puede decir que algo "parece consciente, luego es consciente", a menos que realmente podamos probarlo de alguna manera. Y es muy difícil hacer eso en animales.

A veces se me acusa de negar las emociones de los animales, pero eso no es verdad. Simplemente pienso que metodológicamente es muy difícil probar esto en animales en general.

Pregunta: El Profesor LeDoux también sostiene que lo que realmente es conservado por la evolución no es el comportamiento, como argumentan la mayoría de los científicos y filósofos, sino los circuitos asociados a esos comportamientos que son conservados. Por lo tanto, en su punto de vista, el mismo comportamiento puede variar, pero el circuito particular puede permanecer constante. ¿Podría proporcionarnos un ejemplo apropiado de esta distinción?

Joseph LeDoux: No estoy argumentando que el comportamiento no esté implicado en este proceso evolutivo. Lo que creo es que los comportamientos, en general, están en un nivel diferente de análisis. Por un lado, el sistema nervioso controla el comportamiento,

mientras que por otro lado, la interacción entre el comportamiento del animal y su entorno es controlada por su sistema nervioso.

A medida que las especies evolucionan, cambian y se convierten en otros grupos, otras especies, todas estas transformaciones contribuyen de manera compleja al desarrollo de los comportamientos.

Pero lo que Steven mencionó sobre no ser el comportamiento lo transmitido por la evolución es un punto que hice al pensar en el inicio de la vida. La primera célula que vivió el tiempo suficiente para reproducirse y dar origen a otras células debía ser capaz de detectar y responder al peligro en su entorno.

Lo que esto significaba para una célula que vivió hace 3.7 o 4 mil millones de años es que tenía que satisfacer varias necesidades esenciales para mantenerse viva, como: identificar y alejarse de elementos peligrosos (como toxinas) del ambiente. Pero también debía ser capaz de identificar e incorporar nutrientes, equilibrar fluidos y electrolitos. Y para que la especie continuara, la replicación era esencial.

Si una célula bacteriana primitiva encuentra un alto nivel de acidez, usa sus flagelos para alejarse. De lo contrario, no sobrevivirá el tiempo suficiente para reproducirse. Estos se convirtieron en requisitos fisiológicos fundamentales de la vida, de todas las cosas

que han existido. Sin embargo, cada especie resuelve "sus" problemas a su manera.

Por lo tanto, lo que sostengo es que lo que heredamos a lo largo de cuatro mil millones de años de vida no es el comportamiento en sí, sino la necesidad de responder al peligro, identificar nutrientes, equilibrar fluidos, iones y reproducirse. Estos son requisitos de supervivencia para cualquier organismo vivo, ya sea una sola célula o un organismo gigantesco como nosotros, los seres humanos.

Por lo tanto, no estaba tratando de decir que heredamos nuestra amígdala de las bacterias, lo que heredamos de las bacterias fue la capacidad de sobrevivir, la necesidad de ser capaces de detectar lo que es perjudicial y útil, y permitir cualquier otra cosa que el animal u organismo tenga que hacer para sobrevivir.

Yo llamo a esto "necesidades de supervivencia" o "estrategias de supervivencia" que tienen implementaciones biológicas específicas de manera única para cada grupo de animales y para cada especie de animal, y hasta cierto punto, para cada individuo animal.

En resumen, todos los animales tienen estas necesidades de supervivencia, pero solo los organismos con sistema nervioso poseen circuitos de supervivencia.

Pregunta: Contrario a muchos neurocientíficos, el profesor LeDoux no está de acuerdo en que las emociones estén biológicamente ligadas al cerebro, sino que argumenta que surgen del procesamiento cognitivo inconsciente. ¿Por qué cree que tantos neurocientíficos tienen una visión tan diferente de la suya sobre la naturaleza de las emociones?

Joseph LeDoux: Depende de quién sea el neurocientífico y en qué está interesado. La teoría cognitiva de las emociones no es nueva. William James tenía una versión. Pero la versión moderna se remonta a la teoría de la disonancia cognitiva de Leon Festinger en la década de 1950 y a la teoría cognitiva de la emoción de Stanley Schachter y Jerome Singer en la década de 1960. Basé mi investigación en esto con pacientes de cerebro dividido en la década de 1970.

Mi mentor, Mike Gazzaniga, y yo observamos que cuando el hemisferio derecho de un paciente de cerebro dividido generaba un comportamiento y le preguntábamos al hemisferio izquierdo "por qué lo hizo", inventaba, confabulaba, creando una narrativa para tratar de dar sentido a ese comportamiento. Esto es consistente con la teoría de la disonancia cognitiva de Leon Festinger, que sostiene que cuando hay información discordante, se necesita resolver de alguna manera. Los comportamientos generados por un sistema no consciente, en este caso el hemisferio derecho, podrían ser una fuente de estrés o ansiedad,

ya que todos creemos tener libre albedrío, ya sea que lo tengamos o no, pero creemos que lo tenemos. Si nuestro cuerpo está produciendo comportamientos por los cuales no somos responsables, eso es muy perturbador y tenemos que encontrar una manera de lidiar con ello.

Así, Mike Gazzaniga y yo propusimos la hipótesis de que quizás los sistemas emocionales son aquellos en el cerebro que producen estos comportamientos inconscientemente y que pueden requerir algún tipo de interpretación cognitiva para ajustarlos a la evaluación mental y a nuestro autoesquema sobre quiénes somos.

Por eso me enfoqué en estudios con ratones, porque quería entender comportamientos inconscientes que en los humanos podrían desencadenar interpretaciones cognitivas.

La teoría de la emoción de Schachter y Singer surgió directamente de la "Teoría de la Disonancia Cognitiva" de Festinger. Ha evolucionado significativamente desde la década de 1960 cuando Schachter y Singer la propusieron, pero sigue siendo un enfoque muy válido para entender las emociones. Para mí, es una forma mucho más realista de pensar en nuestras emociones.

Imagina que estás en la cima de una montaña y tu mochila cae por el precipicio. El sol se está poniendo, no tienes comida, agua ni abrigo, de repente, te encuentras en un estado de miedo y ansiedad sobre lo

que te va a pasar. Pero el sistema de la amígdala, o lo que yo llamo el sistema de amenaza, evolucionó como un sistema de defensa frente a depredadores.

No hay ningún depredador en la cima de la montaña que te esté causando miedo. El miedo no es una emoción ligada a un tipo específico de estímulo, es una interpretación. Comienzas a monitorear las señales corporales de tu estómago que indican que tienes poco suministro de energía; al mismo tiempo, empiezas a recibir señales termorreguladoras de que no estás lo suficientemente abrigado.

De toda esta situación, empiezas a preocuparte por la posibilidad de morir de hambre, congelarte o deshidratarte hasta la muerte. Puedes sentir miedo por todas estas razones en la vida que no tienen nada que ver con la presencia de depredadores en tu contexto.

Pregunta: Eso es interesante. ¿Qué tipo de implicaciones clínicas tiene su enfoque en relación, por ejemplo, a los trastornos de ansiedad? Desde un punto de vista teórico, ¿Qué hay de nuevo en su perspectiva?

Joseph LeDoux: Mi grupo de investigación recientemente publicó un artículo en la revista *Molecular Psychiatry* titulado "Colocando el *mental* de

vuelta en *mental*".⁸⁴ La idea es que todo el enfoque de la psicoterapia, el tratamiento de los trastornos mentales que comenzó en la década de 1950, fue impulsado por una agenda conductista que marginó la mente consciente subjetiva.

En la década de 1950, enfoques como la terapia conductual y la psicofarmacología se convirtieron en enfoques estándar. La terapia conductual, obviamente influenciada directamente por Skinner, Watson y otros, abrió el camino hacia la terapia cognitiva, que comenzó con un enfoque más mental (Aaron Beck, el fundador, era psicoanalista), pero gradualmente se volvió más basada en métricas objetivas y algo conductista.

Pero, ¿Qué hay de la psicofarmacología? ¿Quiénes están trabajando en estos laboratorios de la industria psicofarmacéutica, probando animales para descubrir medicamentos que ayuden a las personas? Estos investigadores fueron entrenados por conductistas en las décadas de 1940, 1950 y 1960.

La suposición era que, al poner a una rata en alguna situación amenazante o estresante y darle un medicamento que la haga menos tímida conductualmente, se asumiría que es menos tímida porque es menos temerosa o ansiosa, y que al darle el

⁸⁴ Publicación original: <https://www.nature.com/articles/s41380-021-01395-5>.

medicamento a un ser humano, la persona será menos asustada o ansiosa.

Sin embargo, la industria farmacéutica comenzó a abandonar este enfoque de "anti-ansiedad" porque no generó ningún conocimiento nuevo. La mayoría de las clases de medicamentos para tratar la ansiedad se encontraron accidentalmente en la década de 1960, como los inhibidores de la recaptación de norepinefrina o los benzodiazepinas.

El problema es que, en pocas palabras, un medicamento que hace que las ratas se congelen o eviten el frío no va a cambiar significativamente la angustia humana, la ansiedad, la preocupación y el miedo que una persona experimenta. Necesitamos ayudar al paciente a controlar la fisiología del comportamiento, pero eso no es suficiente, también necesitamos tomar en serio la experiencia subjetiva del paciente si queremos resolver sus problemas.

Pregunta: Dado el carácter innovador de su trabajo y su desviación de las visiones tradicionales sobre las emociones, especialmente al atribuir su origen al procesamiento cognitivo inconsciente en lugar de vínculos biológicos con el cerebro, ¿Ve usted algún paralelo o influencia de las teorías propuestas por Sigmund Freud? Freud también exploró el ámbito de la mente inconsciente y su papel en la formación del

comportamiento humano y las emociones. ¿Reconoce alguna resonancia o diferencia de las ideas freudianas en el desarrollo de sus propias teorías?

Joseph LeDoux: Lo que diría es que los desarrollos que ocurrieron en la década de 1950 para crear nuevas formas de terapia fueron esfuerzos para escapar de Freud, porque se consideraba que su enfoque era demasiado subjetivo. Sin embargo, al deshacerse de la subjetividad que Freud introdujo en la psiquiatría, los científicos de la mente "tiraron al bebé junto con la bañera". En lugar de decir "Bien, tal vez el inconsciente profundo y oscuro no sea toda la historia de la mente y no todo sea represión sexual" y todas las demás críticas a Freud, eso no significa que debemos desechar toda la subjetividad.

Quizás no podamos corregir todo solo cambiando el comportamiento. Como mencioné anteriormente, incluso la "Terapia Cognitivo-Conductual" (TCC) se convirtió en una especie de enfoque basado en el relato del comportamiento. Cuando Beck y Ellis iniciaron la TCC en las décadas de 1950 y 1960, tenía elementos más subjetivos. Sin embargo, estos fueron posteriormente eliminados cuando toda la industria se centró en el pago de seguros. Los terapeutas tenían criterios específicos y métricas objetivas que podían utilizar para identificar o categorizar el problema del paciente.

Si miramos el NIMH RDoC en Inglaterra, por ejemplo, encontramos todas estas métricas que utilizamos para identificar y clasificar aspectos objetivos, y en algún lugar en medio de esa larga lista, encontramos el relato verbal, que es lo que más se acerca a la experiencia subjetiva del paciente. Sin embargo, el relato verbal se trata exactamente igual que cualquier otro síntoma. Es en este detalle donde realmente debemos preguntarnos: ¿Por qué las personas buscan a un terapeuta en primer lugar? Porque se sienten mal y quieren sentirse mejor. Por lo tanto, en mi opinión, debemos corregir cómo se sienten los pacientes, no solo cómo se comportan y eso implica tomar en serio la subjetividad.

Pregunta: Imagina el ejercicio en el que tendrías que programar un nuevo tipo de ser humano desde cero y sus respectivos circuitos de supervivencia. ¿El Profesor LeDoux cree que podríamos programar estos circuitos para que los seres humanos busquen la felicidad, el bienestar o el placer?

Joseph LeDoux: Definitivamente, creo que podríamos utilizar una nueva forma de programación humana, lo cual es una buena idea porque como especie no estamos en un buen camino en este momento. Si planeamos este tipo de proyecto, siempre debemos comenzar con lo positivo y lo negativo e intentar construir algo a partir de ahí. Pero debemos

preguntarnos, en términos de los mecanismos básicos, lo siguiente: Cuando se refuerza a una rata con un comportamiento, ¿Es el placer lo que causa ese refuerzo?

Toda la idea de los centros de placer en el cerebro, desarrollada por Olds y Milner en la década de 1950, surgió por accidente. Olds escribió un artículo científico titulado "Centros de placer del cerebro". Algunos años después, le pregunté a Peter Milner, cuando aún estaba vivo: "¿Qué estaban realmente pensando cuando investigaban estos centros de placer?"

Y él me respondió: "Bueno, en primer lugar, no estábamos estudiando el placer; lo que estábamos tratando de hacer era encontrar mecanismos de refuerzo en el cerebro". Si leemos el artículo que Olds publicó, veremos que la palabra "placer" nunca aparece en el artículo y solo está presente en el título. Quizás los editores de la revista añadieron el título al artículo, y Olds se dejó llevar por él y comenzó a promover la idea de "centros de placer".

Más tarde, Rory Wise comenzó a afirmar que la dopamina es la sustancia química del placer, pero tanto Wise como Olds rechazaron la idea de que estaban estudiando el placer y que la dopamina era la sustancia química del placer.

Pero una vez que "el gato sale de la bolsa" con estas cosas, "nunca vuelve". Ahora, todos se refieren al

refuerzo en términos de recompensa y placer, pero en última instancia, lo que estamos discutiendo son los efectos de la dopamina en la actividad celular que causa un cambio en el comportamiento.

Esto no significa que un animal no esté experimentando placer, o que una persona no esté experimentando placer, pero eso no es lo mismo que el refuerzo de un comportamiento y esta distinción debe ser clara cuando buscamos estudiar estas emociones positivas.

Pregunta: Para concluir, me gustaría saber qué piensa sobre el problema que más ha ocupado a los filósofos. ¿Cree que estamos más cerca de descubrir cómo se forman o se sostienen los “qualia” en el cerebro humano? ¿O aún estamos lejos de poder ofrecer una respuesta sólida al problema difícil de la conciencia?

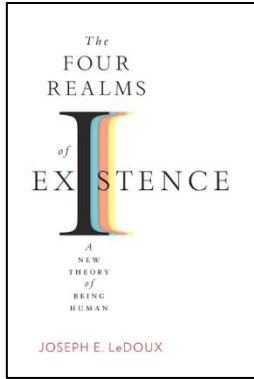
Joseph LeDoux: Una respuesta que puedo ofrecer se basa en el siguiente consejo: Los científicos deben tener cuidado con los conceptos que adoptan de la filosofía.

La filosofía es un sistema de reglas y razonamiento que es útil, pero esto no significa que sea así como el cerebro realmente funciona. No estoy diciendo que los “qualia” no existan, pero creo que el llamado “problema difícil de la conciencia” fue formulado de una manera que nunca podrá resolverse porque presupone el dualismo mente-cuerpo.

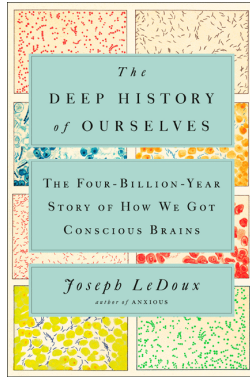
Esto puede ser interesante desde el punto de vista filosófico, pero adoptar una perspectiva dualista sobre el cerebro es un fracaso desde el punto de vista científico. Creo que la mayoría de los neurocientíficos son materialistas y realmente queremos descubrir los mecanismos que producen los "qualia" en la mente.

Tengo un nuevo libro titulado "The Four Realms of Existence: A New Theory of Being", donde propongo una posible manera de pensar sobre los "qualia", así que tal vez sea un buen lugar para que el lector busque más respuestas.

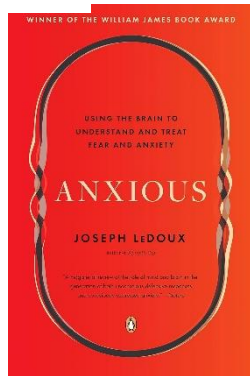
Books by Joseph LeDoux



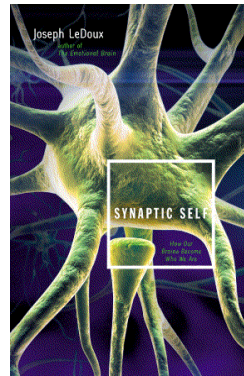
2023



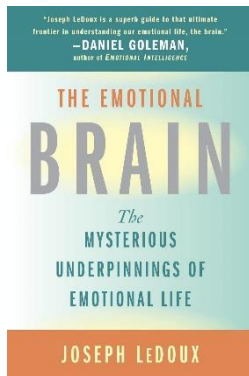
2019



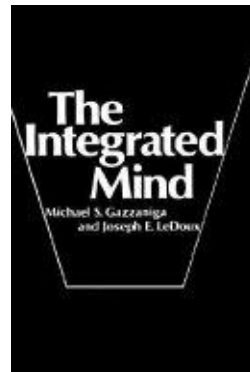
2015



2001



1996



1978

Conclusion

Este libro buscó introducir al lector a algunos de los debates más actuales y fascinantes en filosofía y ciencia de la mente. Así como los valientes navegantes lusitanos que se atrevieron a cruzar el cabo de las "Tormentas", hemos enfrentado las complejidades de los misterios de la mente y los innumerables desafíos de navegación que estos plantean.

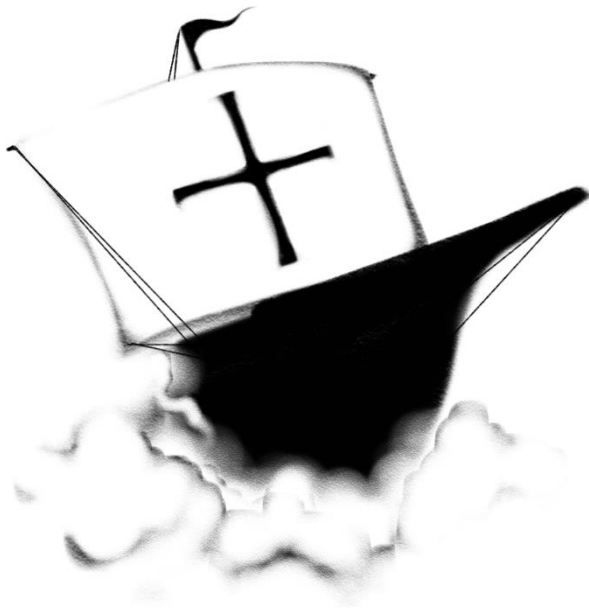


Imagen 26: "Odissea della mente" © Gaia Amoroso

El cabo de las tormentas se encuentra en ciudad del Cabo, Sudáfrica, siendo uno de los puntos más al sur de la tierra. Debido a su posición peculiar, era considerado intransitable debido a las constantes y violentas tormentas y las condiciones de navegación imposibles. Hasta 1488, ningún ser humano había logrado pasar este cabo.

Todo cambió con el navegante portugués Bartolomeu Dias, quien por primera vez logró doblar el cabo de las Tormentas, que luego sería rebautizado por el Rey Juan II como cabo de "Buena Esperanza". Este evento estableció por primera vez una ruta marítima directa entre el océano Atlántico y el océano Índico, conectando Occidente con Oriente, específicamente con la India. Este acontecimiento cambió para siempre la historia económica, cultural y política de nuestro mundo.

Como Bartolomeu Dias, espero que algunos de los avances señalados en este trabajo puedan desarrollarse con mayor robustez en los próximos años, con la esperanza de transformar las "Tormentas" de la mente en un cabo de "Buena Esperanza". Cada página de este libro es como otro día en alta mar, buscando la tierra que quizás nunca llegue. Quizás la curiosidad de cada uno de nosotros sea el viento necesario para impulsar los próximos descubrimientos que puedan emerger de los faros de esta oscuridad.

¿Valió la pena? Confiemos en las palabras ofrecidas por el poeta:

“¿Valió la pena? Todo vale la pena
Si el alma no es pequeña.
Quien quiera pasar más allá del Bojador
Debe pasar más allá del dolor.
Dios al mar dio peligro y abismo,
Pero en él reflejó el cielo.”⁸⁵

Que esta “Odisea de la Mente” brinde al lector un puerto seguro en los puertos del entendimiento y la reflexión. Con el horizonte a la vista, deseo que cada lector encuentre en este libro una brújula confiable para sus propias exploraciones y que la sabiduría adquirida en esta expedición ilumine caminos hacia futuros desconocidos, donde la mente humana pueda explorar más allá de sus fronteras actuales.

Para concluir esta reflexión y el libro, una nota menos poética (una vez más). Este libro surge inspirado en diversos cursos en línea que he organizado como profesor principal en los últimos 2 años. En este

⁸⁵ Fernando Pessoa, Segunda parte: X. Mar portugués del libro *Mensagem*. Versión original en portugués: “Valeu a pena? Tudo vale a pena / Se a alma não é pequena./ Quem quer passar além do Bojador / Tem que passar além da dor. / Deus ao mar o perigo e o abismo deu, / Mas foi nele que espelhou o céu.

contexto, he tenido el privilegio de llegar a más de 600 estudiantes de unos 35 países diferentes.

A todos ellos, mi profundo agradecimiento por haberme enseñado mucho más de lo que posiblemente haya podido transmitir. Además, la aguda curiosidad de los estudiantes, con sus preguntas e intervenciones, ha inspirado muchos debates y ha hecho que los cursos sean mucho más estimulantes y fascinantes para todos, especialmente para mí.

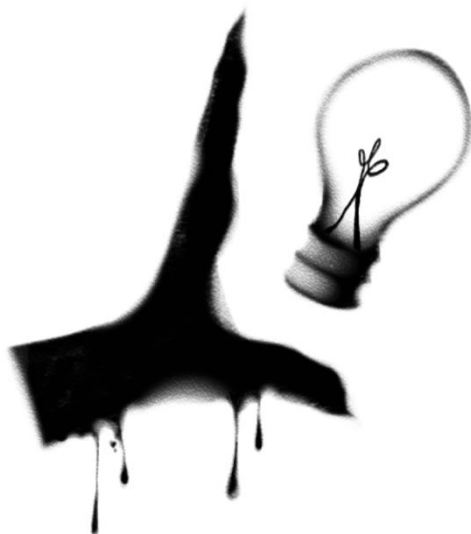


Imagen 27: "La coscienza palpabile" © Gaia Amoroso

¡Que continúe la Odisea de la Mente!

Agradecimientos

Marco Aurelio, filósofo estoico y emperador romano, comienza su célebre obra "Meditaciones" agradeciendo a todas las personas que se cruzaron en su vida y de quienes aprendió algo particular. Ahora bien, una obra de esta magnitud no podría haber ocurrido sin la contribución directa e indirecta de un número variado de personas, por lo que sigo los pasos del emperador, esperando no dejar a nadie fuera.

Comienzo, por supuesto, agradeciendo a los 8 increíbles pensadores que me han otorgado el honor y la confianza de compartir el escenario de este libro: David Chalmers, Susan Blackmore (y Alison Seldon), Sir Roger Penrose (y Helen McGregor), Nicholas Humphrey, Joseph LeDoux, Anil Seth, Christof Koch y Karl Friston, un enorme agradecimiento.

En la ejecución de este libro, Tássia Vianna fue fundamental en las transcripciones de los debates orales en inglés, convirtiéndolos en diálogos escritos que fueron luego aprobados por cada uno de los participantes y traducidos al portugués. Ana Monteiro fue esencial en la creación creativa de la portada de este libro, cuyos rostros fueron dibujados a mano por su puro talento y creatividad. Finalmente, una gratitud

especial a Gaia Amoroso quien, con su talento, hizo que este libro fuera hermoso y no sólo para la mente, sino también para los ojos. A todas, un sincero agradecimiento.

A los colegas de la “Universidad de Oporto”, especialmente a la profesora doctora Sofia Miguens, coordinadora del grupo de investigación “Mind, Language and Action Group” donde realizo mi investigación, que ha sido la mayor influencia para pensar la filosofía de manera abierta al mundo. También al profesor Doctor José Meirinhos, director del “Instituto de Filosofía”, y a la señora directora, profesora doctora Paula Pinto Costa, por apoyar diversos proyectos sin poner ningún obstáculo. A Isabel Marques, gestora de ciencia del Instituto, gracias por toda su paciencia.

A continuación, agradezco a todos aquellos que han contribuido para que mis ideas lleguen al mundo, sirviendo como estímulo constante en mi investigación. En Brasil, a Gabriel Mograbi y Paulo Taddei de la “Universidad Federal de Río de Janeiro”; Nythamar de Oliveira de la “Pontificia Universidad Católica de Porto Alegre”; y Maria Luiza Ilenaco junto con Osvaldo Pessoa Jr. de la “Universidad de São Paulo”.

En Rumanía, a Florin Piscociu y su equipo de Mindlifeline (Léa Chibany, Luana Aldea y los demás miembros) han sido un apoyo constante, cuya ética de trabajo me inspira todos los días. En Malta, Ian Rizzo,

Valdeli Pereira, Francois Zammit y los demás amigos de “Philosophy Sharing Foundation” me han mostrado la importancia de una filosofía pública, desconfiada de la torre de marfil y accesible para todos los ciudadanos. En Chipre, el entusiasmo y el deseo de conocimiento de Panayiotis Stavrou es una motivación constante.

En Chile, la amistad y admiración por el trabajo realizado por Ricardo Ramirez, Camilo Garcia, Marcelo Martín, Nicole Nakousi y otros colegas, quienes me otorgaron el privilegio, a los 30 años, de recibir el título de “Profesor Honorario” de la “Facultad de Medicina Andrés Bello”, un reconocimiento que espero honrar a lo largo de mi vida.

En el contexto de mi proyecto de “Ética de la Inteligencia Artificial en Medicina”, varios colegas deben ser mencionados: Pekka Mäkelä, Raul Hakli y Pii Telakivi (Uni. Helsinki); Radu Uzskai (Uni. Bucarest); Simona Tiribelli (Uni. Macerata); Heidi Mertes (Uni. Gante); Inês Dutra (Uni. Oporto), Sabina Leonelli (Uni. Exeter) y, más recientemente, Antonio Chella y el grupo de “Estudios Avanzados de la Universidad de Palermo”.

A los amigos de la “Universidad Católica de Braga”, con quienes tuve una colaboración breve pero intensa, cuyo enfoque en la multidisciplinariedad fue un aliento: gracias a Augusto Soares Da Silva, Carlos Morujão, Padre João Onofre, Paulo Dias, João Duque, António Melo, Padre Bruno Nobre, Elton Marques, Ângela Leite,

Tânia Oliveira y Padre José Lopes por recibirme en su casa.

Nunca debemos olvidar a quienes vieron algo en nosotros que nadie más vio. Al filósofo Manuel Curado ("Uni. Minho") y al neurocientífico Georg Northoff ("Uni. Ottawa"), supervisores de mi doctorado, gracias por todas las orientaciones en esta aún breve pero energética carrera académica.

A los estudiantes de doctorado a quienes he tenido/tengo el privilegio de orientar y que confiaron en mi guía, gracias a Jaroslav Malík, Romeu Ivoelá y Maria Luiza Ienacco por todas las colaboraciones.

De manera aleatoria, a todos los colegas en las diversas instituciones que han valorado el intercambio de conocimientos: Klaus Gärtner, Robert Clowes, Gloria Andrada, Rui Vieira da Cunha, Sâmara Costa, Sara Fernandes, Bárbara Sousa y Brito, James Grayott, João Cordovil, João de Fernandes Teixeira, Leonel Moura, Ralph Bannell, Roberto Pereira, Diogo Gurgel, Diana Tavares, Dina Mendonça, Joana Rita Sousa, Tomás Magalhães Carneiro, Ângelo Milhano, Björn Lundgren, Jorge Gonçalves, Luísa Neto, Peter Singer, Luiz Meirelles, Noam Chomsky, Yinchun Wang, entre muchos otros cuyos nombres mi memoria no alcanza a recordar pero que siempre están presentes en mi mente.

Para concluir esta ya larga nota de agradecimientos, debo mucho a mucha gente. Marco Aurelio atribuye a los dioses "el hecho de que mi constitución haya sobrevivido tanto tiempo a este tipo de vida". Solo me queda agradecer a mis grandes amigos y familia, quienes me han ayudado en esta "constitución" dedicada al conocimiento: Sr. Joaquim, Maria José y familia (Ana, Manuel, João y Maria Augusta); Rafael, Pedro, João Pedro y Carolina; hermanos Kevin y Dylan; Carlos y Zaida (los verdaderos responsables de esta "constitución"); abuela Arminda, primos Sandra, Nuno, António, Ana y tíos y tías; Mónica y Kiara; Isabel e Isaura; Roberta y Bruno; Ana Maria, Gilberto y Ana Mafalda; a Luna y a los gatos Miró, Tareco, Mia, Didi, Manchinhas y Amarelinho (nombres creativos, lo sé).



Imagen 28: "La relatività del tempo" © Gaia Amoroso

A todos, un profundo agradecimiento.

Biografía del autor

Steven S. Gouveia (Sion, 30/09/1992) es Investigador Asociado en el Grupo de Mente, Lenguaje y Acción del "Instituto de Filosofía" de la "Universidad de Oporto" (Portugal), financiado por la "Fundación para la Ciencia y la Tecnología" (CEECIND.02527.2022), donde lidera un proyecto de 6 años sobre "Ética de la Inteligencia Artificial en Medicina", con colaboraciones de las Universidades de Exeter, Yale y Helsinki (cf. <https://trustaimedicine.weebly.com/>).

Se graduó en Filosofía y a los 22 años comenzó su Doctorado en Filosofía de la Mente en la Universidad de Minho (Portugal), bajo la supervisión del filósofo Manuel Curado, siendo investigador visitante en "Mentes, Imágenes Cerebrales y Neuroética" en el "Instituto Real de Salud Mental", "Universidad de Ottawa" (Canadá), bajo la supervisión del neurocientífico y psiquiatra Georg Northoff.

Tras finalizar su doctorado (2021), fue Investigador en el "Centro de Estudios Filosóficos y Humanísticos de la Universidad Católica Portuguesa", en Braga (Portugal). Desde junio de 2023, es Profesor Honorario en la "Facultad de Medicina" de la "Universidad Andrés Bello", Viña del Mar, Chile, título otorgado coincidiendo con el Premio Nobel de Sir Roger Penrose. También es

consultor ético en la empresa de "Altruismo Eficaz" "Altruistic Careers".

Ha publicado varios libros académicos. En 2016, editó el libro "Filosofía y las Artes en Portugués", invitando a algunos de los artistas más relevantes de la práctica artística nacional, como la artista Joana Vasconcelos, el rapero Valete, el músico Fernando Ribeiro (Moonspell), Leonel Moura (IA), la poeta Ana Luísa Amaral, entre otros.

En 2017, coeditó el libro "Pensando la Democracia" en Portugués, con prólogo de Noam Chomsky, y, con Ana Figueiredo Sol, editó Bioética en el siglo XXI en Portugués. Ese mismo año, también editó "Filosofía de la Mente: Perspectivas Contemporáneas", su primer libro internacional con Manuel Curado por "Cambridge Scholars Publishing".

Al año siguiente, publicó su primer libro como autor en Portugués con editorial Húmus titulado "Reflexiones Filosóficas: Arte, Mente y Justicia", con prólogo del filósofo de la mente João de Fernandes Teixeira.

En 2019, publicó tres libros editados: "Percepción, Cognición y Estética" y "Cine y Filosofía: Superando Divisiones", por el influyente editor Routledge y, para Vernon Press, coeditó, con Manuel Curado, La Película Interna de los "Autómatas: Ciencia y Filosofía de la Mente".

También publicó el libro editado “La Era de la Inteligencia Artificial: una Exploración” por el mismo editor, que cuenta con la participación de algunos de los pensadores y transhumanistas más influyentes, como Daniel Dennett, Ben Goertzel (creador del famoso robot Sofia), David Pearce, Natasha Vita-More, Roman V. Yampolskiy y Vernor Vinge, entre otros.

Al año siguiente, publicó otro libro editado con Bloomsbury, la misma editorial responsable de publicar Harry Potter, titulado “La Filosofía y la Ciencia del Procesamiento Predictivo”.

Además, publicó su libro más buscado hasta la fecha, en Portugués: “Homo Ignarus: Ética Racional para un Mundo Irracional” con editorial Minerva, que cuenta con el prólogo de Peter Singer, un libro centrado en varios problemas éticos como el voto, el humor, la eutanasia o la inteligencia artificial.

En 2022, publicó su tercer libro como autor en inglés por el editor de Nueva York Palgrave, parte del grupo Springer Nature, con el título “Filosofía y Neurociencia: un Análisis Metodológico.”

Todavía en el mismo año, publicó un libro editado titulado “Inteligencia Artificial: Conversaciones sobre el Nuevo Mundo” con la participación de varios académicos como Peter Singer, Wulf Loh o Sabina Leonelli. En 2023, coeditó nuevamente un libro con

Manuel Curado titulado "Mentes Predictivas: Viejos Problemas y Nuevos Desafíos" por editorial Vernon.

Además de esta intensa producción académica, es el presentador y productor del documental internacional "La Era de la Inteligencia Artificial: un Documental", que cuenta con la participación de investigadores internacionales y es accesible de forma gratuita en YouTube, con más de 55,000 vistas.

Ha sido ponente en decenas de conferencias científicas revisadas por pares en Portugal y en el extranjero. Además, ha sido invitado como conferencista en universidades de todo el mundo, en Brasil, Argentina, Chile, Malta, Italia, República Checa, Corea del Sur, Chipre, Rumanía, entre otros. Además, fue ponente en la conferencia de "Ciencia de la Conciencia" en dos ocasiones, la conferencia internacional más relevante sobre estudios de la conciencia. También fue ponente en TEDx NOVA, en septiembre de 2023.

Ha sido solicitado para diversas entrevistas, participando en programas de televisión y radio como "A Minha Geração" (RTP3), "Sociedade Civil" (RTP2), "Linha da Frente" (RTP1), "Much Barulho para Nada" (RTP2), "Filosofia na Rua" (Antena 2), además de varios podcasts, como "45 Graus", "Despolariza", "Pergunta Simples", "Desassossego", "Smart Meat", entre muchos otros.

Es el principal profesor de varios cursos en línea sobre temas como democracia, conciencia o ética, con la participación de profesores invitados como Sir Roger Penrose, Peter Singer, Noam Chomsky, Slavoj Zizek, Paul Bloom, entre otros.

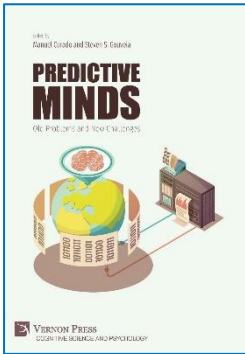
Fue fundador y editor principal de "Apeiron - Student Journal of Philosophy" (2012-2016), que contó con la participación de académicos como Noam Chomsky, Peter Singer, Daniel Dennett o Noël Carroll.

Finalmente, ha sido el principal organizador de varias conferencias internacionales, trayendo a Portugal académicos como Peter Singer (Princeton), Luciano Floridi (Yale), William Child (Oxford), Shaun Gallagher (Memphis), Dan Zahavi (Copenhague), Karl Friston (Londres), David Papineau (Nueva York), Tim Crane (Cambridge), entre otros.

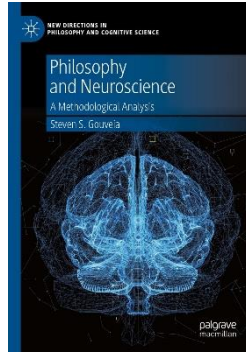
Más información se puede encontrar en: www.stevensgouveia.weebly.com.

Proyecto financiado por el CEEC Individual 2022.02527.CEECIND de la Fundação para la Ciencia y la Tecnología, en el Grupo de Mente, Lenguaje y Acción del Instituto de Filosofía de la Universidad de Oporto (Facultad de Letras, Via Panorâmica s/n, P-4150-564 Oporto, Portugal).

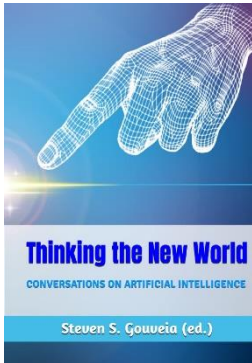
Libros del autor



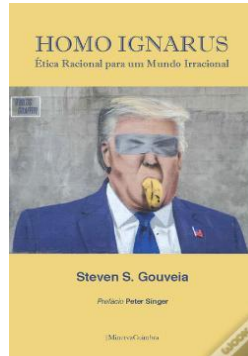
2023



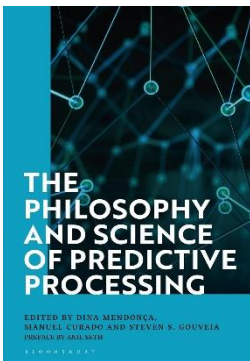
2022



2022



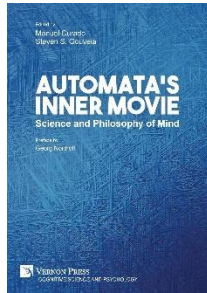
2020



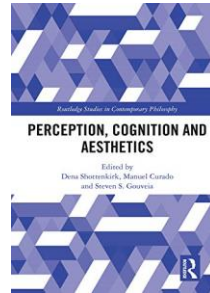
2020



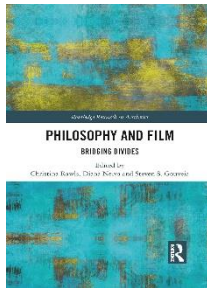
2020



2019



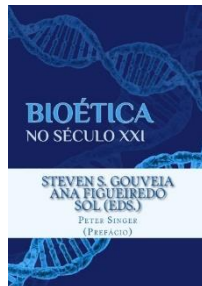
2019



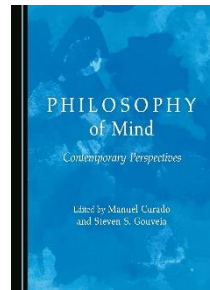
2019



2018



2018



2017



2017



2016