



COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

FACULTAD DE CIENCIAS HUMANAS Y
SOCIALES

Experiencia de libertad: una explicación computacional

Autor/a: Lucía Mollinedo Montero de Espinosa

Director/a: Ignacio Echegoyen Blanco

Madrid

2023/24

Resumen

La consciencia ha sido uno de los objetos de estudio más demandados desde diversas disciplinas, incluyendo la neurociencia, la psicología y la inteligencia artificial. Una de las cuestiones centrales en este campo es entender cómo la consciencia emerge la actividad cerebral y cómo influye en nuestra percepción y comportamiento. Desde una visión computacional de la mente se propone que la consciencia puede ser entendida como un proceso computacional, generado por la compleja interacción de redes neuronales en el cerebro. Además, la experiencia de libertad es uno de los pilares fundamentales de la condición humana. Desde la filosofía hasta la psicología, se ha debatido ampliamente sobre qué significa realmente ser libre y cómo esta experiencia se relaciona con los procesos mentales. El principal objetivo es analizar el papel de la consciencia en la experiencia de libertad; ¿en qué medida la consciencia determina la sensación de control sobre nuestras vidas? Se concluye que es un tema profundamente complejo y que la relación entre ellas e incluso la explicación del funcionamiento de la consciencia está sujeta a interpretaciones variadas y a menudo contradictorias.

Palabras clave: consciencia, libertad, computación, mente y cerebro.

Abstract

Consciousness has been one of the most attractive objects of study across various disciplines, including neuroscience, psychology, and artificial intelligence. One of the central issues in this field is understanding how consciousness emerges from brain activity and how it influences our perception and behavior. From a computational view of the mind, consciousness is proposed to be understood as a computational process, generated by the complex interaction of neural networks in the brain. The experience of freedom is one of the fundamental pillars of the human condition. From philosophy to psychology, there has been extensive debate about what it truly means to be free and how this experience relates to mental processes. The primary objective is to analyze the role of consciousness in the experience of freedom; to what extent does consciousness determine the feeling of control over our lives? It is concluded that this is a deeply complex issue, and the relationship between them, and even the explanation of consciousness' functioning, is subject to varied and often contradictory interpretations.

Key words: consciousness, freedom, computation, mind, brain.

| | |
|---|-----------|
| INTRODUCCIÓN | 1 |
| FINALIDAD Y MOTIVOS | 1 |
| ESTADO DE LA CUESTIÓN..... | 1 |
| OBJETIVOS Y PREGUNTA DE INVESTIGACIÓN | 2 |
| INTRODUCCIÓN A LA CONCEPCIÓN DE LA MENTE Y LA CONSCIENCIA | 3 |
| ¿POR QUÉ ES IMPORTANTE EXPLICAR LA MENTE Y LA CONSCIENCIA? | 3 |
| DUALISMO | 5 |
| MONISMO | 7 |
| CONSCIENCIA..... | 10 |
| DEFINICIÓN DE CONSCIENCIA | 10 |
| ANÁLISIS Y OBJETO DE ESTUDIO | 12 |
| <i>Conceptos de consciencia.....</i> | <i>12</i> |
| <i>Niveles de consciencia.....</i> | <i>14</i> |
| <i>Medidas de consciencia.....</i> | <i>15</i> |
| POSICIONES DE CONSCIENCIA..... | 17 |
| ¿PUEDE UNA MÁQUINA TENER CONSCIENCIA? | 19 |
| CONCLUSIÓN Y LIBERTAD | 22 |
| BIBLIOGRAFÍA..... | 25 |

Introducción

Finalidad y motivos

En el complejo campo de la filosofía, pocas cuestiones han fascinado y desafiado tanto a la humanidad como la naturaleza de la consciencia, la mente y la libertad. Estos temas están intrínsecamente entrelazados y han sido objeto de reflexión y debate desde tiempos inmemoriales, impulsando a filósofos, científicos y pensadores de todas las épocas a adentrarse en los abismos de la experiencia humana.

Precisamente, la propuesta de este trabajo de investigación es explorar esos límites difusos de la consciencia y la libertad. ¿Qué es la consciencia, esa experiencia que define nuestra existencia y nos permite percibir y comprender el mundo que nos rodea? ¿Cómo se relaciona con la mente, un complejo entramado de procesos cognitivos, emocionales y percepciones que dan forma a nuestra identidad y realidad interna? Y quizás, ¿en qué medida nuestra consciencia y nuestra mente influyen en nuestra capacidad para ejercer esa libertad?

A través de un análisis de las teorías más influyentes y de los hallazgos más recientes desde la filosofía hasta la neurociencia, el propósito de este trabajo es proporcionar una perspectiva que pueda aportar algo de claridad a las grandes incógnitas que actualmente dominan muchos ámbitos de la comunidad científica. Ante las grandes preguntas, se ha intentado a lo largo de este trabajo abordar cuestiones que todavía no están resueltas pero que, al analizarlas en profundidad, esclarecen el misterio de la consciencia.

Estado de la cuestión

Searle (2002) considera que la consciencia es el problema más importante de las ciencias biológicas. Plantea que, durante mucho tiempo, el tema de la consciencia ha sido abordado de una manera “distinta” a cualquier otro problema científico, ante la diversidad de opiniones de los científicos que, sin duda, han obstaculizado la investigación. Las dos opiniones más generales a cerca de esta cuestión son: los que consideran que la consciencia es un programa computacional, un software en el hardware del cerebro; y, los que creen que la consciencia es únicamente un método de procesamiento de la información.

Hay muchas preguntas que rodean a la consciencia: ¿cómo son los procesos cerebrales que causan los estados conscientes? ¿podemos explicar la consciencia con la teoría existente o necesitamos conceptos nuevos y revolucionarios para explicarla? ¿cuál es el nivel apropiado

para explicar la consciencia? ¿las neuronas y la sinapsis? Es decir, la consciencia, a pesar de ser un problema biológico, tiene una serie de cuestiones filosóficas que hace que la complejidad de la cuestión aumente y se convierta en algo mucho más difícil de abordar.

Searle (2002) cree que una de las grandes cuestiones es la falta de consenso en la definición de consciencia y, por tanto, el mal uso del concepto en los distintos contextos. A pesar de que, según Giménez – Amaya et al. (2007), la neurociencia haya experimentado un desarrollo importante en las últimas décadas el problema mente – cerebro y las relaciones entre ellos sigue vigente ahora no solo en el campo de la filosofía sino como objeto de estudio de la ciencia.

Por último, en la era de la inteligencia artificial, la comprensión de la consciencia, sus límites y su relación con la libertad se ha vuelto mucho más compleja. La emergencia de sistemas inteligentes capaces de realizar tareas que antes se consideraban exclusivas de la mente humana plantea interrogantes fundamentales sobre la naturaleza misma de la consciencia ¿Puede una máquina ser consciente? ¿Hasta qué punto la inteligencia artificial puede simular o incluso emular la experiencia subjetiva? Además, el creciente uso de algoritmos y sistemas autónomos plantea desafíos éticos y filosóficos en torno a la libertad individual y la toma de decisiones. En este trabajo, se propone investigar como la inteligencia artificial ha ampliado el panorama de la comprensión de la consciencia y cómo enfrentamos los desafíos que plantea en términos de definición, límites y repercusiones en nuestra concepción de libertad.

Objetivos y pregunta de investigación

Los objetivos generales de este trabajo son:

- Hacer una revisión bibliográfica de las distintas teorías de la mente.
- Estudiar la consciencia en profundidad.
- Analizar el pensamiento de los distintos autores con respecto a la consciencia y la libertad.

Los objetivos específicos son:

- Explorar la experiencia subjetiva de libertad que tienen los seres humanos con respecto a las máquinas.
- Investigar los límites de la consciencia y las teorías contemporáneas sobre ella.

Introducción a la Concepción de la Mente y la Consciencia

Los novedosos métodos de la Neurociencia moderna y sus resultados relevantes demuestran con claridad lo mucho que todavía queda por saber e investigar acerca del funcionamiento del cerebro. No obstante, han supuesto un impulso decisivo para volver a plantear el clásico problema de las relaciones entre la mente y el cerebro. Quizá lo más característico es que el problema parece haber dejado de ser un monopolio de la filosofía y se ha convertido en inevitable para la misma ciencia (Gimenez – Amaya y Murillo, 2007).

Como se expondrá en el presente trabajo, hay muchas teorías que abarcan la relación entre la mente y la consciencia e incluso la mente y el cerebro, pero la falta de claridad y consenso entre ellas hace que la solución sea vista como algo muy lejano e incluso imposible de alcanzar. Resulta muy paradójico que el ser humano busque estudiar la consciencia utilizando su propia consciencia. Sin embargo, la única manera de estudiar la consciencia es presuponiendo que tenemos una e incluso en algunos casos, asumiendo que otro tipo de organismos poseen una.

¿Por qué es Importante Explicar la Mente y la Consciencia?

La consciencia humana es una de las grandes incógnitas del mundo actual; está caracterizada por ser uno de los temas que más confusión y frustración generan en los grandes pensadores. ¿Cómo es posible que, uniendo varios cientos de miles de células, aparentemente simples (neuronas) interconectadas entre sí, yo pueda estar escribiendo este texto, o dos personas puedan tener una conversación, escuchar música, apreciar la belleza de un paisaje, saborear su comida favorita o incluso, reflexionar sobre su propia existencia? Es decir, ¿cómo puede la materia volverse sobre sí misma y saber que existe?

Antes de empezar a definir la consciencia, lo óptimo sería establecer la diferencia entre la mente y el cerebro. El cerebro es el centro biológico que recibe los estímulos del medio interno y externo, transduciendo a señales nerviosas e integrándolas en una misma experiencia unificada. Por mente, entendemos el procesamiento de esa información e integración de componentes cognitivos, emocionales y sensoriales, tanto conscientes como inconscientes. El problema mente-cerebro consiste en el hecho de que poseemos experiencia subjetiva y podemos dar cuenta de ella y, además, somos capaces de examinar los órganos mentales involucrados en ella de manera científica, pero la unidad entre ambos es difícil de encontrar (Giménez-Amaya y Murillo, 2007). Autores como Nagel (1974), creen que, si no existiese la

consciencia, el problema mente – cerebro sería muy poco interesante, pero el hecho de que exista lo convierte en algo muy difícil de resolver.

La gran incógnita tiene muchas maneras de manifestarse. Por ejemplo, Dennet (1995), habla de que ajustamos inconscientemente nuestra conducta, atención y movimientos oculares para sincronizar la música que podemos estar escuchando en un momento dado, con la imagen que estamos viendo a través de nuestros ojos y si llegamos a plantearnos y a pensar sobre esa sincronización, resulta hasta más difícil de modelar que lo que realmente está ocurriendo dentro de nosotros.

Otra cuestión que es necesario abordar es la posibilidad de naturalizar la conciencia, y si esa posibilidad existe ¿de algún modo cambiaría la imagen que tiene el hombre de sí mismo? Alan Turing, por su parte, desarrolló el funcionamiento y procesos del primer ordenador y concluyó que el mismo era un cerebro electrónico (Cañedo-Argüelles, 2015).

En cuanto a la conciencia, aunque no existe definición consensuada sobre este fenómeno, lo más sencillo resultaría definirlo desde una concepción “natural”: Cañedo-Argüelles (2015) afirma que la conciencia es “una dimensión autotransparente de la vida psíquica, en virtud de la cual el sujeto pensante se convierte en espectador activo de sí mismo”. No obstante, siguiendo esta definición asumiríamos la posibilidad de autoconsciencia. Por tanto, se ha propuesto otra definición aún más neutral: todo aquello de lo que el ser humano puede dar cuenta (Cañedo-Argüelles, 2015).

La importancia del estudio de la mente recae sobre la necesidad de responder al problema de naturalizar la mente y la consciencia (estudiarla científicamente), para lo que aún no se ha hallado respuesta alguna. La primera cuestión consiste en el hecho de si es posible atribuir patrones neuronales concretos a la función de la consciencia, porque está claro que no todas las áreas cerebrales están implicadas en la producción de experiencias conscientes (Doty, 1998). La segunda se refiere a la complejidad de la percepción sensorial y la naturaleza—desconocida y multifacética—de las experiencias que crean los órganos sensoriales. En la actualidad, no podemos describir la variedad de sensaciones que experimentamos a través de la interacción de las neuronas. La tercera cuestión se centra en el movimiento muscular, la parte consciente consiste en fijar la meta (ej. Coger un bolígrafo), el problema debe abordarse en cuanto a la naturaleza del impulso que pone en marcha cualquier programa de este tipo. La cuarta cuestión, explicada extensamente por Sperry al descubrir que existe una consciencia humana de manera independiente en cada hemisferio (Doty, 1998), pero la gran incógnita es, como se ajustan para producir una única experiencia propia de cada ser humano. El quinto y último problema hace hincapié en la aparente ilimitada y duradera capacidad de la memoria;

¿cómo puede el sistema neuronal, incluso desde su inmensurable complejidad, seguir almacenando a lo largo de toda la vida numerosos segmentos detallados de la información que fluye hacia él? (Doty, 1998)

Debido a la falta de consenso presentada entre los distintos autores que han intentado dar respuesta a esta pregunta de una manera filosófica y científica, han surgido distintas corrientes que han luchado por abordar el tema de la mente y la conciencia. En los siguientes apartados, profundizaré sobre algunas de ellas para analizar la evolución de los distintos autores y la concepción de la mente y la conciencia desde una visión filosófica y computacional, para intentar dar respuesta a la experiencia subjetiva de libertad.

Dualismo

El dualismo se basa en la idea de que la mente no es lo mismo que el cerebro por estar compuesta de una sustancia que la hace especial, no de la materia ordinaria que forman los órganos físicos de nuestro cuerpo. Conviene introducir el pensamiento de Descartes y la explicación dualista que ofrece sobre el cuerpo y el “alma” (a lo que hoy en día conocemos como “mente”): Descartes (1972) consideraba que el cuerpo y el alma eran dos sustancias creadas y diferentes, y con esencias distintas; la esencia del alma es el *pensamiento* (“res cogitans”) y la del cuerpo, la *extensión* (“res extensa”). Mientras que el cuerpo puede ejecutar infinidad de acciones sin intervención del alma, ésta, como sustancia pensante puede regular el cuerpo (aunque no siempre). Hasta dónde llega la unión de la mente sobre el cuerpo escapa de nuestra posibilidad de conocimiento y debemos limitarnos a creer en su mutua interacción (Wozniak, 1992).

Pero, precisamente esto, era la debilidad de Descartes y de los futuros objetores de su teoría, la incapacidad de explicar esa interacción: los órganos sensoriales deberían informar a la mente (enviar percepciones, datos o ideas) para que posteriormente ésta pueda reflexionar sobre ellas e informar al cuerpo para dirigir sus acciones y que sean apropiadas a la respuesta que se quiere dar. Es por esto por lo que, al dualismo de Descartes también se le ha denominado interaccionismo cartesiano o dualismo interaccionista. Descartes creía que el punto de encuentro entre la mente y el cerebro era la glándula pineal y, de hecho, la “dibujaba” en sus libros con medidas exageradas. Sin embargo, la percepción consciente solo se produce cuando el cerebro transmite el mensaje a la mente y, el cuerpo solo emite respuestas en cuanto la mente le transmite las órdenes pertinentes. Pero ¿cómo se transmite la información desde la glándula pineal a la mente? El propio Descartes no pudo responder a esta pregunta porque carecemos de

información suficiente para definir de qué sustancia está hecha la mente y, por tanto, tampoco sabemos cómo esta sustancia se ve afectada por los ingredientes físicos que provienen del cerebro (Dennet, 1995).

Según Searle (2007), el dualismo no fue tomado en serio por los filósofos y científicos, a pesar de que el público general lo acogiese como modo de entender el funcionamiento del mundo. Defiende que la consciencia no es ontológicamente reducible a procesos neuronales (es decir, no puede limitarse a explicarse en términos de sus componentes fundamentales o entidades más básicas) y, además, cuestiona que el dualismo sea una forma de irreductibilidad ontológica de la consciencia. Cree que no podemos explicar la realidad de manera coherente a través del dualismo, porque desde él, nuestros estados conscientes no se cohesionan con una explicación del mundo real. El dualismo, como he explicado anteriormente, postula dos dominios distintos: cuerpo y alma. Pero, la interacción entre ellos es inexplicable, precisamente porque explicar cómo los procesos cerebrales en un dominio ontológico (cuerpo) producen la consciencia en otro dominio ontológico, es imposible. Y, esto trae una consecuencia a la que se ha denominado “epifenomenalismo”; si lo que caracteriza a la consciencia son las cualidades de subjetividad, unidad, intencionalidad y ausencia de materialidad o pertenencia al mundo físico, ¿cómo se explica que participe y funcione en la realidad física si son características que no están presentes en la realidad?

Como he mencionado, Descartes admite la existencia de cuerpo y alma, cerebro y mente, dos elementos. Pero la pregunta, según Cañedo – Arguelles (2015) recae sobre la diferenciación y límites de esas dos sustancias, y la resuelve diciendo que todo aquello que se pueda explicar de manera mecánica corresponde a la *res extensa* (cuerpo) y todo lo que sea incapaz de explicarse mecanizadamente es *res cogitans*, alma.

Por otro lado, si asumimos la existencia de dos sustancias como hacía Descartes, debemos intentar entender la interacción entre ellas. A pesar de que, en su momento, el hecho de establecer conexiones entre la *res cogitans* y la *res extensa* fuese acogido por toda la comunidad de pensadores filósofos y científicos, a día de hoy se considera casi imposible. Lo que defendía Descartes era que el corazón genera y comunica a la sangre la energía que el organismo necesita. A su vez, del cerebro se extrae un líquido al que llama “espíritus animales” y que, canalizado por los nervios hasta los músculos, hace que se muevan las articulaciones. Siguiendo esta teoría, nos volveríamos a encontrar en un bucle de preguntas al plantearnos: ¿quién dirige el líquido? ¿quién lo destila? Descartes responde a estas cuestiones de una manera muy simple: el líquido merodea por el cerebro y sale casi por donde puede, de la manera más natural posible, hacia las estructuras. Este líquido, responde al “soplo del espíritu”, la glándula

pineal. Cuando el alma quiere algo con mucha fuerza, la glándula se mueve de la manera exacta para cumplir la voluntad del alma (Dennet, 1995).

En este sentido y como iré desarrollando más adelante, entra en juego la voluntad del individuo. Descartes creía que era de dos tipos: la que busca satisfacer las necesidades de alma (amar a Dios o aplicar nuestro pensamiento a algo no material) y las acciones que se cumplen en nuestro cuerpo (caminar, correr, comer, etc.). Pero, aun separando los dos tipos de acciones, la *res extensa* está unida al cuerpo y no se puede separar de ella, pero tampoco se puede atribuir a alguna de las partes del cuerpo porque el alma es de una naturaleza que carece de extensión y de las propiedades de la materia de la que está compuesta el cuerpo. Esto lo demuestra defendiendo que cuando algún órgano o parte del cuerpo se deteriora, todo el cuerpo se deteriora, mientras que el alma permanece intacta, aunque ocurra esto. Cuando el cuerpo muere, el alma se separa de él (Descartes, 1972).

Descartes dejó un legado al que muchos científicos y pensadores se vieron obligados a responder: aunque la mayoría asumen la existencia de la mente y el cuerpo o hacen alguna distinción entre ellas, no todos acogen la relación causal que Descartes defiende. Malebranche, por ejemplo, cree que no hay influencia de la mente sobre el cuerpo o viceversa, sino que Dios es la única causa operativa y verdadera. Sin embargo, defender la predominancia de la mente sobre el cuerpo también es afirmar la existencia del cuerpo y lo material y en cuanto se establece una distinción entre los dos aparece el problema mente – cuerpo y su relación (Descartes, 1972).

Monismo

Comenzaremos estableciendo la diferencia entre un monismo “de tipo” (“*type monism*”), que defiende que sólo hay un tipo de cosa presente y concreta existente en el mundo y un monismo más idealista, en el que autores como Bradley, creen que solo existe lo absoluto (“*the absolute*”), un ser humano individual y absoluto. El monismo fisicalista—que desarrollaré más adelante—rompe con la teoría de Descartes y el dualismo cartesiano, que pretende conservar la independencia de lo mental y lo físico. No obstante, conserva la idea de Descartes de que el cuerpo y la mente son inseparables, interaccionan, porque confirma la identidad de ambos, se compromete con la idea de que no pueden ser comprendidos y que su unión es ininteligible. Gale Strawson (citado en Weekes, 2012) coincide con Berkeley y cree que el problema de la distinción entre la mente y el cuerpo es muy complejo y está incomodando la posibilidad de pensamiento de la comunidad científica, lo que debemos de

hacer para avanzar y progresar en esta cuestión es no asumir una distinción real entre la mente y el cerebro. La pregunta es si esta afirmación es lógica o simplemente por conveniencia para su estudio.

Platón fue el primer filósofo que dio al cerebro un lugar prioritario en el *cosmos*; en su diálogo *Timeo*, describe el cosmos como algo viviente, trascendental, creado por un ser divino. El cerebro fue lo primero que creó el demiurgo (creador del cosmos), para que los seres humanos pudiesen percibir el mundo a través de los ojos, oídos y razonar a cerca de él (Giordano, 2011).

Whitehead (citado en Weekes, 2012), insiste en que la ontología de la naturaleza debía ser monista, pero con esta afirmación no incluía a la mente como objeto de estudio o, por el contrario, lo consideraba algo físico. Lo que ocurre es que, si asumimos una concepción monista de la realidad considerando que la mente inunda al mundo físico, caeríamos en lo que conocemos como *panpsiquismo*. Si asumimos un pansiquismo (contrario al materialismo que defiende que la materia es lo único que existe) como concepción de la realidad, ante la pregunta de: ¿de qué están hechas las cosas?, una visión dualista desde el pansiquismo hablaría de una distribución mental de la “sustancia mental” (*“mental stuff”*) en el mundo físico. Y, una perspectiva monista, considera que toda la materia de la que están hechas las cosas es puramente sustancia mental (*“mental stuff”*) y, sólo sustancia mental.

Resulta conveniente establecer la diferencia entre el monismo “idealista” y el monismo “materialista”. El primero de ellos considera que el universo es puramente mental o espiritual. No existe nada más que las ideas y cualquier fenómeno mental es considerado también una idea. El materialismo considera que la realidad existente es material, todos los estados mentales son físicos. El problema recae sobre la necesidad de formular una teoría materialista de la mente sin negar la existencia de los estados conscientes. En esta misma línea, el “funcionalismo” se plantea la función de las cosas y asume la existencia de mente y cerebro realizando una comparativa con la distinción entre hardware-software de un ordenador. Busca entender la diferencia de una manera conceptual, más que real, al contrario que los pensadores más reduccionistas. El monismo reduccionista, rechazando una de las dos mitades de Descartes, no rechaza la dicotomía. Precisamente porque han preservado la idea de una distinción real entre lo mental y lo físico, se permiten tener la opción de desechar una y preservar la otra. Según Searle (2004) los estados mentales se definen como estados que tienen determinadas funciones y la función se explica en términos de una relación causal entre un estímulo externo y un estado mental. Por ejemplo, María percibe que llueve y causa en ella una creencia de que llueve; a su vez, esta creencia y el deseo de no mojarse, causan el comportamiento de coger un paraguas.

Dentro del materialismo encontramos el conductismo metodológico defendiendo que la mente es el comportamiento del cuerpo y no hay nada más allá de eso. El conductismo metodológico descubre las leyes del estímulo – respuesta y convierte la psicología en una psicología del comportamiento, más que de la mente. El conductismo lógico establece una teoría aún más fuerte, partiendo de una crítica al dualismo de Descartes (que explicaré más adelante). La afirmación de un estado mental, por ejemplo, una creencia de que va a llover o un dolor en la pierna significa lo mismo que una serie de creencias sobre el comportamiento actual o posible de esa persona. Por ejemplo, María cree que va a llover significa lo mismo que el hecho de que María guarde las sillas de la terraza dentro cierre las ventanas, etc. Es decir, la idea de tener un estado mental es el hecho de estar dispuesto a determinados modos de comportarse. Por tanto, esto debería analizarse en términos hipotéticos: “si p, luego q”. Es decir, vendría a ser una reducción de lo mental a operaciones lógicas, tales como el comportamiento (Searle, 2004).

Por otro lado, dentro del materialismo también encontramos el fisicalismo y la teoría de la identidad. Para los fisicalistas, lo que consideramos “mente” es un cerebro y a lo que llamamos “estados mentales” son estados del cerebro y del sistema nervioso. La teoría de la Identidad nace de la necesidad de refutar al conductismo. Defienden que los estados mentales son idénticos a los estados cerebrales, de la misma manera que el agua es lo mismo que el H₂O, que un relámpago es lo mismo que un impulso eléctrico y que un dolor físico es lo mismo que un estado mental. No obstante, ante el nacimiento de esta teoría surgieron objeciones de muchos tipos. La principal fue la crítica a la falta de aclaración con respecto a las propiedades de las cosas. Es decir, si se afirma el hecho de que el agua es lo mismo que el H₂O, se deben identificar las propiedades del agua y las propiedades del H₂O y posteriormente, concluir que son lo mismo. Porque, de lo contrario, suponiendo que son propiedades independientes, estaríamos ante dos cosas muy distintas: emergen propiedades de la mente y propiedades de la materia. Por tanto, para que la Teoría de la Identidad tenga sentido, debería convertirse en una teoría dualista. Es decir, si todos los estados mentales son estados cerebrales, debe haber algunos estados que son mentales y otros que no lo son. Los estados mentales tienen propiedades mentales y los otros, propiedades físicas, cerebrales. Esto es un tipo de dualismo (Searle, 2004).

Consciencia

Hace no mucho tiempo el tema de la consciencia generaba muy poco interés entre los psicólogos, neurocientíficos y filósofos. En general, se ha considerado como un tema que no merece investigación neurocientífica, para eso ya “se estudiaba el cerebro” y era concebido como algo muy subjetivo y personal como para analizarlo en profundidad. Además, la investigación biológica de la consciencia ha generado mucha contienda entre la sociedad: ¿Cuáles son los procesos cerebrales implicados en los estados conscientes?, ¿cómo se crean esos estados en las estructuras cerebrales? A día de hoy, muchas preguntas carecen de respuesta exacta o correcta y no son más que planteamientos que por el momento, “satisfacen” las necesidades de los investigadores, o eso se cree (Searle, 2002).

Definición de Consciencia

Antiguamente, se hablaba de “tener consciencia” o ser “consciente” en relación a la capacidad de lenguaje, ser inteligente o mostrar algún tipo de conducta o comportamiento determinado. Pero la consciencia no depende del comportamiento de cada uno, como ha quedado demostrado en las personas que sufren una parálisis total del cuerpo y, aun así, sus niveles de consciencia son estables; o incluso cuando dormimos y soñamos, tenemos consciencia. Por tanto, esa afirmación supondría asumir que todas aquellas personas que han perdido la capacidad de lenguaje, bebés o animales carecen de consciencia (Seth, 2021)

Otro problema se plantea en cuanto se asume que la consciencia es todo aquello de lo que podemos dar cuenta. Es decir, una persona es consciente del árbol que está viendo, el bolígrafo que está cogiendo o la canción que está escuchando, pero también es consciente sin percibir sensorialmente aquello de lo que damos cuenta. A este fenómeno lo denominamos consciencia transitiva, cuando tenemos un pensamiento sobre algo que no percibimos directamente o pensamos en la sensación que nos genera sentir algo (por ejemplo, pensar en la suavidad de la seda) (Rosenthal, 2009).

El último fenómeno es controvertido y está siendo actualmente estudiado por la comunidad científica. Como se ha mencionado previamente, somos conscientes de aquellas cosas que percibimos o pensamos, pero a su vez, esos pensamientos y percepciones, pueden ser conscientes o no conscientes, como en el caso de la *percepción subliminal* (Rosenthal, 2009). Es decir, hay determinados estímulos que pueden ser percibidos de manera inconsciente y tener un efecto medible objetivamente en el comportamiento del sujeto. Se han realizado experimentos en los que reduciendo de manera progresiva la intensidad de un estímulo

presentado (hasta el punto de que sea casi imperceptible para el sujeto) demuestre haberlo percibido en la realización de distintas tareas (Núñez, 2012).

Para una criatura consciente, existe una sensación real de sí mismo, es decir, una experiencia subjetiva del “yo”, del “tú” e incluso de “vaca” o “caballo”. Para aclarar esta idea, es necesario hacer referencia a la fenomenología: las propiedades subjetivas de la experiencia consciente. Es decir, por qué la experiencia subjetiva de ver algo, tiene unas propiedades subjetivas (forma, estructura y cualidades) concretas y distintas a la experiencia subjetiva de sentir algo o tocar algo. A estas propiedades las llamamos *qualia*; es decir, quizá lo más difícil de entender es porqué y cómo la actividad neuronal genera experiencias subjetivas conscientes como la rojez del color rojo o lo doloroso de un dolor como un golpe en el dedo meñique del pie. Lo interesante es que muchos neurocientíficos son capaces de explicar cómo cuando vemos una manzana roja, la luz impacta en la retina y se transforma en impulsos eléctricos que van del tálamo al córtex y genera en nosotros el color rojo. Pero según Chalmers (citado en Kanai y Tsuchiya, 2012), el “hard problem” es el carácter subjetivo de esa experiencia, es decir, los *qualia* solo existen desde dentro porque ¿cómo explicamos o demostramos que somos conscientes de la rojez del rojo? Los *qualia* no pueden ser demostrados de una manera objetiva ni medible. Sin embargo, para los seres humanos la existencia de los *qualia* de hambre, dolor y percepción son la esencia de nuestra existencia; es decir, es lo más íntimo que conocemos, y lo más misterioso para estudiar. Algunos autores defendían que los *qualia* no tenían propiedades, alegaban que era un término filosófico que lo único que conseguía era generar confusión a la hora de explicar la consciencia. Pero Dennet, estableció las propiedades de: inefables, intrínsecos y privados. Una vez establecidas las propiedades, se cuestionaron las funciones de los *qualia*; Richard Gregory (citado en Kanai y Tsuchiya, 2012) defendía que los *qualia* cumplían una función para el presente, servían para establecer que la percepción es de aquí y ahora, para separar lo que pasaba en el mundo exterior de lo que pasaba dentro de nuestra mente.

Se ha defendido la consciencia como “cualquier experiencia subjetiva”. Sin embargo, es muy simple, probablemente por el hecho de que es un fenómeno muy complejo y, por tanto, entendido de manera muy incompleta (Seth, 2021). De hecho, Chalmers (1997), cree que es lo más familiar que existe y a la vez, lo más misterioso. Y afirma que el término consciencia es imposible de definir y mucho menos explicar lo que hace o cómo ha evolucionado, excepto en términos que son ininteligibles sin saber lo que significa la consciencia. Propone ofrecer ejemplos e ilustraciones que caracterizan a la consciencia pero que, en ningún caso, la definen. No obstante, coincide con la idea anterior y aclara que un estado mental se caracteriza como

consciente en el momento en el que tiene un sentimiento cualitativo (“*qualitive feel*”), es decir cualidades fenomenológicas o *qualia* (Chalmers, 1997). En contraposición a esto, Dennet (1995) intenta definir y explicar la consciencia del ser humano desde la fenomenología y habla de “jardín fenomenológico” para referirse a los elementos que componen nuestra mente y así poder llegar a la naturaleza última de las cosas. Lo divide en tres partes: experiencias del mundo exterior (imágenes, sonidos, olores...) experiencias del mundo interno (visiones, sonidos de nuestros sueños, conversaciones con nosotros mismos...) y experiencias emotivas, estas son complejas porque encontramos, por un lado, dolores, cosquillas, sensaciones de hambre, etc.; y, por otro lado, arrebatos emocionales como la ira, vergüenza, felicidad, rabia...

Análisis y Objeto de Estudio

Conceptos de consciencia.

Sebastián (2022) considera de gran importancia definir con claridad el uso que le damos al término “consciencia” en las diferentes situaciones, además, analiza la diferencia entre la consciencia de acceso como el hecho de “darnos cuenta” y la consciencia fenoménica. Lo primero que resulta conveniente aclarar es la diferencia entre “ser” consciente y “estar” consciente (estados mentales o conscientes). En este sentido, Sebastián (2022) coincide con Rosenthal (2009), afirmando que lo primero es una propiedad que otorgamos o negamos a distintos organismos y lo segundo, es exclusivo de los estados mentales. En esta línea, se habla de que alguien es consciente cuando es capaz de percibir un estímulo y sentirlo, por ejemplo, Pedro es consciente del daño que le ha hecho Juan al pegarle un bofetón. Afirmar esto, sería afirmar que todo aquel que “sienta” es consciente. Pero esto hace aún más grande la incógnita sobre dónde establecer los límites entre criaturas conscientes y no conscientes, por ejemplo, las plantas responden a estímulos del ambiente e incluso algunas se abren o cierran con solo tocarlas, pero ¿realmente alguien se atreve a afirmar que una margarita es un “ser consciente”?

Como desarrollaré más adelante, para intentar resolver esta cuestión se han establecido diferencias entre una criatura “consciente” y una criatura “despierta” en función de los grados de alerta en los que se presente su organismo. Desde la filosofía, se ha hablado de “consciencia fenoménica” que según Rosenthal (2009) es completamente independiente de la “consciencia de acceso” y hace referencia al hecho de que una criatura es consciente sólo si se siente de alguna manera, es decir, *es como algo* ser esa criatura. En relación a lo anterior, la consciencia fenoménica implica que la criatura sea capaz de responder a estímulos del ambiente, pero no

necesariamente tiene estar despierta, como en los sueños. Por ejemplo, ser fenoménicamente consciente implica que se siente *como algo* ser yo y se siente *como algo* (distinto) ser tú.

Por otro lado, se ha realizado una distinción entre criaturas conscientes, que es la de conciencia transitiva e intransitiva. La primera hace referencia al hecho de que somos conscientes del perfume de nuestra madre o de la música de un concierto, mientras que la conciencia intransitiva se refiere también a la conciencia fenoménica, en el sentido de que se refiere al uso de la conciencia cuando nos preguntamos si otra persona es o no consciente. Está relacionado con la conciencia fenoménica porque solo podemos preguntarnos esto cuando *se siente de alguna manera ser nuestro compañero de trabajo*.

Resulta muy conveniente para Sebastián (2022) utilizar el término “darse cuenta” para referirse a “ser consciente”. Al ver una película tenemos una experiencia visual consciente, nos damos cuenta de que estamos viendo la película. Y es precisamente esto a lo que se refiere la “conciencia de acceso” previamente mencionada, al “darse cuenta”. De hecho, según Rosenthal (2009) este tipo de conciencia juega un papel fundamental en las teorías de conciencia como espacio de trabajo global (GWT) que he explicado previamente. Precisamente por el hecho de que un estado es consciente si tiene un efecto global en los distintos procesos cognitivos del individuo (memoria, comportamiento, etc.), es decir, si se ha “dado cuenta” de ellos.

Recapitulando, la conciencia fenoménica suscita mucho interés en la comunidad científica, precisamente por la dificultad que genera ofrecer una explicación sobre ella y sobre cómo establecer su lugar en la naturaleza. Es uno de los motivos por los que Chalmers (1996), lo sitúa en el corazón del problema mente – cuerpo y, ejemplifica esta conciencia ofreciendo un símil con los “zombis”. Imaginemos un zombi como los de las películas, sí percibe lo que tiene en frente, es consciente de lo que ve y oye y, por tanto, sí que tiene conciencia transitiva. Además, está despierto, pero para él *no se siente como nada* ser un zombi, por tanto, carece de experiencias fenoménicamente conscientes, por muy despierto que esté. Es decir, lo fundamental sería el papel causal de la conciencia fenoménica.

Por otro lado, la conciencia de acceso parece mucho más simple de entender. Si estamos viendo una película, somos conscientes de que vemos la película, de los colores, de la banda sonora y los personajes. También podemos para la película y volverla a poner y, además, tengo una experiencia fenoménicamente consciente, *se siente como algo* para mí ver esa película. En este sentido, nuestro cerebro genera una experiencia consciente de la película y lo pone a disposición de otros sistemas como el sistema de formación de creencias o actitudes o de los mecanismos de acción. Por tanto, tiene sentido pensar que una de las funciones de la conciencia podría ser esta; seleccionar determinados contenidos para entregárselos al resto de

sistemas, que sean representados, razonados y dar una respuesta acorde con las necesidades del entorno. Por tanto, siguiendo esta teorías, la consciencia de acceso tendría un papel más funcional, pero sería distinta de la consciencia fenoménica.

Niveles de consciencia.

Como hemos visto, son muchos los autores que se han preocupado por la conciencia, los contenidos conscientes, su naturaleza, sus grados y niveles. Como en todo, hay discrepancias sobre cuáles son los contenidos de toda la información que percibimos que acceden a nuestra consciencia creando las experiencias y contenidos conscientes. Es decir, ¿qué diferencia aquello de lo que somos consciente de aquello de lo que no?

En este sentido, Searle (2002) defiende que hay habilidades y conocimientos generales que posibilitan el funcionamiento de nuestros estados mentales; a esto él lo llama “Trasfondo”. La premisa fundamental de esta tesis es que los fenómenos intencionales operan únicamente dentro de un conjunto de capacidades de Trasfondo que no poseen intencionalidad por sí mismas. Un concepto clave para entender la existencia de estos fenómenos es el de “Red”, la existencia de fenómenos mentales requiere formar parte de una Red que incluya otros fenómenos intencionales. Esto se sostiene porque resulta imposible pensar que los estados intencionales determinen de una manera aislada sus condiciones de satisfacción. La posesión de creencias o deseos implica la necesidad de una red completa que abarque otras creencias y deseos. Y, la relación entre Red y Trasfondo radica en que, además de la Red, es esencial postular un Trasfondo de capacidades que no forman parte de la Red. La totalidad de la Red necesita un Trasfondo, ya que los elementos de la Red no se auto – interpretan ni se auto – aplican.

La cuestión es, diferenciar entre Red y Trasfondo. La Red es la parte del Trasfondo que es capaz de causar intencionalidad consciente. En el cerebro hay millones de neuronas operando que causan estados conscientes e inconscientes, los estados conscientes tienen toda la gama de colores y variedades de lo que percibimos desde el exterior. Hay incluso otras neuronas que solo se utilizan para fijar la creación de esos estados conscientes. Por tanto, parece que habría que hacer una distinción entre lo intencional y lo no intencional (Bächler, 2005)

Según Nuñez (2012) si la mente seleccionase para crear contenidos conscientes toda la información que recibimos del exterior, nos acabaríamos saturando, a pesar de ser una máquina muy compleja preparada para operar con muchísima información a la vez. Por eso, aquellos contenidos que pasan a consciencia son seleccionados de manera rigurosa. Y, cuando hablamos

de “niveles de consciencia” nos referimos a aquella información que es seleccionada más o menos intensamente, completa o duradera. Por tanto, en este sentido, no habría niveles como tal, sino un único nivel donde se recoge toda la información y en el que se seleccionan algunos contenidos o no para que formen parte de nuestra actividad consciente. En esta misma línea, Rosenthal (2009) defiende que, por ejemplo, en el caso de la percepción subliminal (que previamente he explicado) percibimos de una manera más genuina y, por tanto, nos “percatamos” desde un segundo nivel.

Lo primero que debemos hacer es diferenciar cuando la consciencia es pasiva (los datos aparecen en consciencia sin más, como cuando vemos una película, somos meros observadores de lo que ocurre) de cuando es activa (elaboración consciente de manera esforzada como cuando estamos estudiando). Por tanto, la “actividad consciente” propiamente dicha se identificaría, según Núñez (2012) con un nivel superior y el resto de información que no es manipulada directamente para crear actividad consciente y, por tanto, nuestra experiencia consciente, como en el caso de la percepción subliminal, correspondería a niveles inferiores.

Chalmers (citado en Bächler, 2005) cree que la consciencia es una propiedad de los seres humanos que existe en la naturaleza y que, por esto, aun cuando se presente en diferentes grados variables en las diferentes criaturas, prácticamente todas presentarían consciencia. No obstante, el propio Chalmers afirma que esto es un hecho casi imposible de aceptar, porque poseemos lo que llamamos “autoconsciencia” y eso nos diferencia del resto.

Medidas de consciencia.

¿Qué significa medir la consciencia?, ¿cómo de consciente estoy siendo ahora mismo escribiendo este texto? ¿Cómo de consciente eres tú al leerlo?, ¿cuál es la diferencia entre ser consciente del todo y ser un trozo de carne que no tiene experiencia de nada?

Según Seth (2021) muchos filósofos hacen referencia a un “termómetro de la consciencia” que es capaz de determinar si una persona, animal o incluso una máquina es o no consciente. Si resulta que la consciencia es algo parecido a la temperatura corporal (medible) sería increíble porque no solo determinaría si algo es o no consciente, sino que incluso podríamos hablar de niveles y grados de consciencia.

Medir la consciencia (aunque no sea como medir la temperatura corporal) resulta muy importante y esencial para controlar la naturaleza de la experiencia subjetiva. La medición convierte lo subjetivo en objetivo, lo cualitativo en cuantitativo, lo vago en preciso.

Medir tiene otro tipo de implicaciones, por ejemplo, la anestesia se suministra a más de un millón de personas cada día, el hecho de poder estabilizar a un paciente por debajo del nivel de arousal correspondiente, temporalmente y sin que sufra daños refleja el hecho de que un “metro” preciso y confiable sería muy valioso. Además, el hecho de crear una medida de consciencia ayuda a esclarecer las incógnitas en el ámbito médico sobre si después de determinados estados conscientes permanecen después de daños cerebrales cuando algunos pacientes son declarados “vegetales” o “inconscientes”.

No obstante, medir la consciencia no significa lo mismo que declarar a alguien como “despierto” o “dormido”. El arousal psicológico no es lo mismo que los niveles de consciencia a pesar de que en múltiples ocasiones correlacionen. Cuando sueñas, estás dormido, pero también hay varias experiencias conscientes en el sueño. Para establecer la diferencia, deberíamos preguntarnos qué significa en el cerebro ser consciente en oposición a estar despierto; no parece que tenga que ver con el nivel de neuronas involucradas. Incluso, atendiendo a los niveles de actividad neuronal el cerebro no parece estar mucho más activo en estados conscientes que en inconscientes. De hecho, se ha demostrado que el ser consciente depende de la manera en la que distintas partes del cerebro se comunican entre ellas más que del cerebro como conjunto (Seth, 2021)

El método más nuevo para medir la consciencia y distinguirla del mero hecho de estar despierto consiste precisamente en esto, identificar y cuantificar las interacciones del cerebro en estados consciente para determinar cómo de consciente está. Massimini (citado en Seth, 2021) testó cómo se comunicaban diferentes partes del córtex mediante la simulación de actividad en una de las zonas y midió como esa actividad recorría las distintas regiones corticales en el espacio – tiempo. Combinaron el electroencefalograma (EEG) con la simulación magnética transcraneal (SMT). La SMT permite al investigador insertar directamente en el cerebro un impulso fuerte y corto de energía, mientras que el EEG registra la respuesta del cerebro a este impulso.

A pesar de que no sentimos directamente la energía que se inserta mediante la SMT, Massimini encontró que la diferencia en la respuesta eléctrica del cerebro ante ellos podría establecer los distintos niveles de consciencia. En niveles más inconscientes, como cuando dormimos, la respuesta es muy simple y muere rápido; mientras que en niveles conscientes la respuesta recorre todo el sistema cortical, desapareciendo y apareciendo, formando patrones complejos. La complejidad de estos patrones demuestra que las distintas partes del cerebro se comunican entre ellas durante estados conscientes y mucho menos durante estados inconscientes. Lo interesante del trabajo de Massimini, es que la respuesta del cerebro durante

los estados conscientes se puede cuantificar mediante un algoritmo que se traduce en un índice de complejidad (PCI). Este índice ha permitido obtener numerosos avances en medicina como el hecho de distinguir entre estados vegetativos y estados mínimamente conscientes, estableciendo de esta manera la diferencia entre estar “despierto” y estar “consciente”.

No obstante, a día de hoy no hay una medida universal para medir el grado de consciencia de una persona. A diferencia de otros aspectos psicológicos y fisiológicos que pueden evaluarse mediante instrumentos de medida específicos, tanto cualitativamente como cuantitativamente. La naturaleza subjetiva e intrínsecamente personal al igual que la falta de consenso, estudio y profundización del fenómeno de la experiencia consciente complica su cuantificación. Aunque existen, como he ejemplificado, diversos enfoques e intentos que funcionan como instrumentos de medida en muchos casos, la carencia de métodos directos para medir la consciencia ha generado un desafío en la investigación científica. Este vacío en los métodos de medida resalta la complejidad única de la consciencia.

Posiciones de Consciencia

La falta de consenso en torno a una teoría de la mente o una explicación neuroanatómica y funcional de la consciencia conduce a los investigadores a cuestionar y revisar constantemente las teorías existentes, lo que a su vez genera discrepancias y debates en la comunidad científica. Esta situación dificulta el avance científico al obstaculizar la acumulación de conocimiento y la consolidación de modelos teóricos sólidos.

Kaní y Fujisawa (2023) crearon un concepto denominado “universality” o universalidad para nombrar aquellas teorías que determinaban si un sistema era consciente, independientemente de su origen o composición (tanto un cerebro humano como un ordenador). Esto significa que una teoría debe ser aplicable a cualquier sistema físico. El término “*universalidad*” se utiliza mucho en física para definir aquellas leyes de la naturaleza fundamentales que permanecen consistentes y se aplican de igual manera en todas partes del universo.

En las primeras décadas tempranas, el estudio científico de la consciencia se centró sobre todo en los correlatos neuronales de la consciencia o “*neural correlates of consciousness*” (NCCs). Esto ha sido relativamente útil para establecer un lenguaje y una metodología común entre los investigadores aún con diferentes posturas, porque básicamente consiste en indagar sobre los estados y procesos del cerebro que están más relacionados con la consciencia. No obstante, tiene muchas limitaciones porque no resulta fácil distinguir los verdaderos “NCCs”

de los requisitos neuronales y consecuencias de la consciencia. Por tanto, en respuesta a estas limitaciones se desarrollaron las Teorías de la Consciencia (ToCS). Algunos consideran que una teoría de la consciencia satisfactoria podría resolver el “hueco” que tanto ha preocupado a la comunidad científica; la duda realmente está en si ese “hueco” algún día será resuelto del todo. Por tanto, las teorías en las que centraré mi revisión en este apartado son tres: (HOTs) Teorías de orden superior (*Higher – order theories* o *HOTs*), Teorías de la consciencia como espacio de trabajo global (*Global workspace theories* o *GWTs*) y Teorías de la Información Integrada (*Integrated Information theory* o *IIT*) (Seth, 2022)

Teorías de orden superior (HOTs)

El aspecto central que une a todas las HOTs es que un estado mental es consciente en virtud de ser el objetivo de un tipo específico de estado meta – representacional. Es decir, representaciones que tienen como objetivo, otras representaciones; el hecho de tener la experiencia visual de estar viendo a un pájaro volar es una meta – representación. La cuestión de estas teorías es que la consciencia involucra correlatos neuronales de un nivel “superior” que aprenden a redescibir esas representaciones que se han decodificado en niveles inferiores de tal manera que pasa a ser una “meta – representación” (Seth, 2022).

Las HOTs se centran en explicar por qué algunos contenidos son conscientes y otros no. Uno de los puntos más polémicos de estas teorías es la intuición que los contenidos de la experiencia perceptiva sobrepasan la información disponible en el “primer orden” como se afirma que ocurre en el contexto de la visión periférica. La propuesta que hacen estas teorías es que los niveles superiores hacen una especie de representación errónea de la fenomenología de la visión periférica en los estados de primer orden.

Estas teorías pueden explicar por qué hay algunos estados que son incapaces de ser conscientes (no pueden ser objeto de una meta – representación) y otros son necesariamente conscientes (deben ser meta – representados). Además, una pregunta que genera mucha intriga es el hecho de si este tipo de teorías explican el carácter fenomenológico de distintas experiencias. Es decir, ¿por qué el carácter fenomenológico asociado a un dolor de rodilla es tan distinto del que acompaña a ver un atardecer? La respuesta desde estas teorías es que el carácter fenomenológico de las distintas experiencias depende de las propiedades que el estado meta – representacional relevante le atribuya. En última instancia, cualquier versión reductiva del enfoque debe explicar por qué la representación de las distintas propiedades de una experiencia genera la fenomenología que genera y cómo esas propiedades se representan gracias a la actividad neuronal (Seth, 2022).

Teorías de la consciencia como espacio de trabajo global (GWT)

Según Kani y Fujisawa (2023) establecen que estas teorías defienden que la consciencia nace de la transmisión de información a través de distintos canales neuronales en el cerebro. Básicamente establecen que aquella información consciente es la que es accesible globalmente para los distintos procesos cognitivos como la atención, evaluación o memoria; a esta información la llamamos “espacio de trabajo global”. Es decir, la experiencia consciente está constituida por la información que es accesible para los sistemas cognitivos de un individuo.

Al igual que las anteriores, esta teoría se centra en qué hace que una representación sea consciente a través de la comprensión del cerebro humano. No obstante, sus principios, pueden no extenderse fácilmente a sistemas no biológicos o a sistemas biológicos radicalmente diferentes. Y, en este contexto, resulta difícil identificar si un sistema de inteligencia artificial posee un espacio de trabajo global.

Teoría de la información integrada (IIT)

Esta teoría se aproxima a la caracterización de la fenomenología desde una visión matemática. Comienza proponiendo unos axiomas sobre el carácter fenomenológico de la experiencia consciente (propiedades que son consideradas verdaderas y aplicables a las formas posibles de conciencia). A partir de esos axiomas derivan afirmaciones sobre las propiedades que cualquier sustrato físico de la conciencia debe cumplir. Por tanto, los sistemas físicos que posean estas propiedades necesariamente poseerán conciencia; Propone que la conciencia debe entenderse en términos de causa – efecto (Seth, 2023).

En principio, esta teoría sería aplicable a cualquier sistema dinámico. Establece el límite entre las partes de un sistema que corresponder a una entidad consciente de aquellas pertenecientes a otras entidades o al medio ambiente (Kani y Fujisawa, 2023).

¿Puede una máquina tener conciencia?

Como he expuesto anteriormente, durante muchos años filósofos, científicos y pensadores han debatido sobre una cuestión que a día de hoy sigue abierta. Tras haber explicado las distintas teorías de la consciencia vigentes en la actualidad y haber entendido (hasta donde se puede) cómo funciona, podemos darnos cuenta de que si realmente no es más que un entramado de neuronas estableciendo comunicaciones e intercambios ¿se podría replicar y hacer una máquina o un robot capaz de superar a los humanos?

Ante esta pregunta, muchos autores han ido planteando sus teorías sobre el funcionamiento de la mente; a continuación, profundizaremos en el debate planteado entre la

teoría del hardware y software de Searle dos de las más destacadas son las de Searle (2004) y Signorelli (2018) por ser las más controvertidas.

Searle (2004) cree que la manera de funcionar del cerebro es como un ordenador digital, así funciona el sistema: a lo que llamamos “mente” es un programa de ordenador o una combinación de programas y los “estados de la mente” son estados computacionales del cerebro. Por tanto, la mente es el programa (software) y el “cerebro” sería el hardware del sistema.

Para entenderlo mejor, aclara que un algoritmo es un método de resolución de problemas mediante la consecución de una serie de pasos, que deben ser finitos y que, llevados a cabo de una manera correcta, la solución al problema queda garantizada. Algo así como realizar una suma en matemáticas (añadir números de manera consecutiva nos dará un resultado acertado). Por otro lado, una máquina de Turing es un sistema hipotético que permite analizar los límites y funcionamiento de un algoritmo y de algunos sistemas lógicos. Es un dispositivo que va leyendo, secuencialmente, los elementos escritos en una cinta infinita. Cuenta con un número finito de reglas, que operan sobre cada elemento. De forma que puede “probar” (hipotéticamente) universalmente todo un conjunto de algoritmos y sistemas lógicos. En ese sentido, se considera un precursor del ordenador.

El Test de Turing se creó para comprobar cuando una máquina se está comportando de una manera inteligente y cuando no. La principal idea del Test de Turing era ver si la máquina era capaz de “engañar” a un ser humano; es decir, ¿puede la máquina actuar de tal manera que un experto pueda no distinguir su actuación de una actuación humana? Por ejemplo, si una máquina responde a preguntas formuladas en chino igual de bien que un hablante nativo de chino, de manera que otros hablantes nativos no puedan distinguir la diferencia entre la máquina y un hablante nativo, entonces tendríamos que decir que la máquina comprende el chino. Es decir, el Test de Turing, expresa una especie de conductismo y, por tanto, la prueba de comportamiento es concluyente para la presencia de estados mentales.

El sentido de estas ideas era pararse a pensar si el cerebro funcionaba, por ejemplo, como una máquina de Turing. Adoptar esta idea como filosofía de vida resultaba muy atractivo y excitante por el hecho de que resolvía los problemas filosóficos sobre el funcionamiento de la mente. Searle (2007) concluye que el cerebro funciona exclusivamente como un ordenador digital y más concretamente como una máquina de Turing. El cerebro, sería un programa o una comunicación de programas y para entender la cognición y el comportamiento humano solo sería necesario averiguar qué programas exactamente se implementan cuando se activan las funciones de la mente como la percepción o la memoria. Además, por el hecho de ser máquinas

de Turing, entenderemos la cognición mediante la reducción de operaciones complejas en operaciones más simples, manipulando ceros y unos.

El cerebro es una entidad física igual que los ordenadores e incluso funciona con señales eléctricas, resuelve problemas complejos y procesa la información, tal y como hacen ellos. Además, la duda aumenta proporcionalmente al desconocimiento del cerebro y la mente; carecemos de la suficiente información para explicar cómo procesa el cerebro todos los datos (a lo mejor no es comparable a un procesamiento computacional). De lo que sí que podemos dar cuenta es que el cerebro funciona con propiedades complejas como las experiencias subjetivas (por ejemplo, emociones, conciencia y comportamiento condicionado). Estas propiedades no tienen una correlación directa con una computabilidad alta, de hecho, el cerebelo es el centro de las conexiones, posee más neuronas que cualquier otra parte del cuerpo y, aun así, parece no estar implicado en la percepción consciente. Por tanto, resulta paradójico que se asimile al cerebro humano con un ordenador por el funcionamiento de las billones de células que operan a una velocidad inimaginable cuando el 80% de estas células se encuentran en el cerebelo que no tiene un papel importante en la percepción consciente (Signorelli, 2018)

Signorelli (2018) considera que el cerebro no debe de ser comparable a un ordenador digital, en contraposición a lo que relataba Searle (2004) ni con su “software” ni con su “hardware” porque esta analogía nubla las propiedades físicas del cerebro. Lo que ocurre, es que muchas veces la neurociencia y la ciencia cognitiva utilizan términos como “computación”, “información”, “procesamiento” para intentar explicar a la comunidad algo tan complejo como el funcionamiento de la mente y la conciencia y a la vez, asumen que no se va a malinterpretar el uso de este tipo de palabras con el funcionamiento de un hardware. X ofrece un ejemplo muy apropiado para romper con la analogía cerebro – ordenador. Cree que es mucho más oportuno comparar la mente con un tambor, porque responde con diferentes vibraciones muy complejas. De tal manera que los “inputs” (golpe) y los outputs (sonido y vibración).

Para que una máquina tenga conciencia o funcione de una manera muy parecida a la manera de trabajar de nuestro cerebro sería necesario que tuviese una serie de capacidades que solo los seres humanos tenemos y que debieran imitar o superar. La primera de ellas es la autonomía, que actualmente y como observamos en el sistema de IA Chat GPT que funciona de manera autónoma buscando información según las preguntas y respuestas que le hacemos e incluso siendo capaz de reutilizarlas para dar una respuesta más satisfactoria a los usuarios. Por otro lado, la reproducción, algo que biológicamente resulta imposible de imaginar, pero si pensamos en robots capaces de arreglarse entre ellos y fabricarse unos a otros dejamos de verlo como algo lejano. No obstante, resulta menos obvio cuando hablamos del pensamiento crítico

y moral si la inteligencia artificial no integra algo que es exclusivo y esencial de los seres humanos: la consciencia (Signorelli, 2018).

Además, para que un sistema tenga consciencia debe tener las dos características fundamentales que acompañan al hecho de poseer consciencia: autoconsciencia y capacidad de “darse cuenta”, por tanto, no podemos imaginarnos una máquina o un ordenador capaz de poseer las capacidades de un ser humano, comparar la mente con un sistema computacional o incluso pensar que podemos replicarlo sin considerar la opción de crear máquinas con consciencia, algo que a día de hoy es imposible porque significaría desarrollar máquinas con un alto nivel de cognición, esto es, sentimiento de confianza, imaginación mental, subjetividad, autoconsciencia, etc. No obstante, el hecho de plantearse replicar la mente humana o comparar el funcionamiento del cerebro al de un ordenador, a pesar de ser imposible, resulta de gran utilidad en términos académicos, éticos y experimentales (Signorelli, 2018).

Nagel (1974) considera que es muy probable que algún día llegemos a entender como el cerebro y la mente origina la consciencia. Pero esto ocasionaría una revolución en nuestra manera de pensar y entender la realidad, el mundo y la ciencia. Porque nuestro sistema no es capaz de concebir lo subjetivo, cualitativo y las experiencias del fenómeno neuronal.

Resulta interesante mencionar que Dung (2023) diseña un test para identificar la presencia de consciencia (que no la ausencia) en un sistema de inteligencia artificial. Con la firme idea de que el hecho de que la máquina pasase el test significaría la presencia de consciencia en el sistema de inteligencia artificial, incluso independientemente de que la máquina estuviese exclusivamente diseñada para pasar ese test. No obstante, incluso ella admite que el test tiene algunas limitaciones que, a mi parecer, son suficientes para descartar la presencia de consciencia (como la entendemos en los seres humanos) en los sistemas de inteligencia artificial.

Conclusión y libertad

Como he mencionado anteriormente, la consciencia guarda una estrecha relación con el grado de libertad que los seres humanos experimentamos. Una vez más vuelven a surgir preguntas sobre si la consciencia toma posesión de todos nuestros actos, por ejemplo, del hecho de coger un bolígrafo y levantarlo ¿Dónde empieza la intencionalidad? ¿en el dedo o en nuestra mente?

Para comenzar a hablar de esto, deberíamos aclarar el término *libertad* para ello, al ser un concepto tan amplio y subjetivo que abarca muchas cuestiones, resulta más claro hablar de

lo que no es la libertad; según Seth (X) no es una intervención del flujo físico del cerebro, tampoco se reduce simplemente a las relaciones causa – efecto o incluso al determinismo, entendido como que todos los eventos están determinados por causas físicas existentes en el universo.

El concepto de libertad, y más profundamente el “sentirse libre” implica una primera sensación de que estoy haciendo aquello que quiero hacer, que a su vez no solo implica valores, deseos, estados psicológicos sino tiempo y oportunidades. No obstante, muchas veces esas acciones—que son voluntarias—como querer comer algo y prepararlo, valga la redundancia, voluntariamente, pueden no serlo porque la sensación de hambre es involuntaria. Para diferenciarlo, cuando percibimos una acción como voluntaria, y por tanto, libre, viene alineada con pensamientos y deseos desligados del mundo o incluso de nuestro cuerpo.

Si hablamos de “grados de libertad” es porque, un sistema es tan libre como deseos satisfaga o acciones realice. Es decir, una piedra no tiene libertad, pero un autobús tiene un grado de libertad: moverse hacia delante o hacia atrás. La voluntad depende de la capacidad de controlar estos grados de libertad, y esta capacidad está integrada en el cerebro mediante millones de conexiones. El neurocientífico Patrick Haggard (citado en Seth X) explica como estas conexiones tienen tres estadios: un “qué” para determinar la acción que vamos a realizar, un “cuando” para establecer el momento en el que se va a realizar dicha acción y un “por si acaso” que permite cancelar o inhibir la acción en el cerebro en el último momento. Dentro del *qué* incluiríamos deseos, apetencias, valores, pensamientos, metas, etc. Que una vez más, están controladas por millones de regiones del cerebro, especialmente de la corteza prefrontal. El *cuándo* se asociaría más a la necesidad subjetiva de moverse, la región motora del cerebro genera una necesidad subjetiva de moverse que nos impulsa a actuar. Por último, el *por si acaso* es una última comprobación de la acción determinada para verificar que esa acción debe ser llevada a cabo.

Pero entonces, ¿la libertad es una ilusión? Esta pregunta ha generado mucho revuelo en la comunidad científica y especialmente al tratarse acompañado del complejo tema de la consciencia, pues bien, una vez más, la respuesta es: depende. Es cierto que, como he explicado, la fenomenología de la voluntad es una especie de alucinación de control autorrealizadora relacionada con tipos específicos de acciones que surgen desde adentro. Al tocar el piano, preparar un café o mientras escribo este texto, estas acciones voluntarias se llevan a cabo con una automacidad y fluidez que traspasan la intuición de que de alguna manera yo causo esas acciones cuando realmente, la mayor parte del tiempo nuestras acciones voluntarias y pensamientos simplemente “suceden”.

Desde otra perspectiva completamente opuesta, el libre albedrío no es una ilusión en absoluto, cada uno de nosotros tiene una capacidad de ejecutar e inhibir acciones voluntarias gracias a la capacidad de nuestro cerebro de controlar esos estados de libertad. Sin embargo, esta libertad es la que se utiliza (desde esta perspectiva) para actuar según nuestras creencias, valores y metas, para hacer lo que queremos y tomar decisiones de acuerdo con quienes somos, más que una libertad que sigue las leyes de la naturaleza o del tejido causal del universo. Es decir, no se podría infravalorar el poder causal que proyectamos sobre las experiencias volitivas, de la misma manera que proyectamos el color rojo en nuestra percepción de una superficie. Además, estas experiencias volitivas resultan indispensables para nuestra supervivencia, provocan acciones voluntarias y sin ellas no lograríamos desenvolvernos en entornos complejos ni aprender de nuestras propias acciones.

Las grandes incógnitas siguen abiertas ¿seremos capaces de explicar el grado de libertad que poseemos numéricamente? ¿cómo continuarán influyendo las nuevas tecnologías en el estudio de la consciencia? ¿cuál es el papel del inconsciente en la experiencia de libertad?

Bibliografía

- Arana Cañedo-Argüelles, J. (2015). *La conciencia inexplicada*. Biblioteca Nueva, S.L.
- Bächler, R. (2005). El problema de la Conciencia. *Ciencia Cognitiva (1)2*. Pp. 161-172.
- Chalmers, J.D. (1996). *The Conscious Mind: In search of a fundamental Theory*. Oxford University Press <https://books.google.es/books?hl=es&lr=&id=0fZZQHOfdAAC&oi=fnd&pg=PR11&ots=qovX4iBQTR&sig=VNRnaBdfyzZJ3mP9gdFJyPcQDqQ#v=onepage&q&f=false>
- Dennet, D. (1991). *La conciencia explicada*. Paidós Básica.
- Descartes, R. (1972). *Las pasiones del alma*. Edaf.
- Doty, R.W., (1998). The five misteries of the Mind and their Consequences. *Neuropsychologia, 36*. Pp. 1069 -1076.
- Dung, L. (2023). Los tests de consciencia en animales son los tests de consciencia en máquinas. *Original Research*.
- Gimenez Amaya, J.M. y Murillo, J.I. (2007). Mente y cerebro en la Neurociencia contemporánea: una aproximación a su estudio interdisciplinar. *Scripta Theologica 39*. Pp. 607-636.
- Giordano, M.(2011). Cerebro y mente en el Siglo XXI. *Revista Digital Universitaria (12)3*. Pp. 1067-6079.
- Kanai, R. y Fujisawa, I. (2023). Towards a Universal Theory of Consciousness.
- Nagel, T. (1974). What Is It Like to Be a Bat?
- Núñez, J.P. (2012). *La mente: la última frontera*. Universidad Pontificia Comillas.
- Roosenthal, D.M. (2009). Concepts and Definitions of Consciousness. *Elsevier Inc*.
- Ryota, K., Naotsugu, T. (2012). Qualia. *Current Biology, 22(10)*.
- Searle, J.R. (2002). La conciencia. *Consciousness and Lenguaje*.
- Searle, J.R. (2007). Dualism revisited. *Journal of Psychology*. Pp. 169-178.
- Sebastián, M.A. (2022). *El problema de la consciencia. Una introducción crítica a la discusión filosófica actual*. Cátedra Teorema.
- Seth, A. (2021). Being You: A new Science of Consciousness.
- Seth, A. y Bayne, T. (2022). Theories of consciousness. *Nature Reviews (23)*. Pp. 439-452.
- Signorelli, C.M. (2018). ¿Pueden los ordenadores tener consciencia y superar a los seres humanos? *Hypothesis and Theory (5)121*.

- Weekes, A. (2012). The Mind-Body problem and Whitehead's Non-reductive Monism. *Journal of Consciousness Studies* (19)9. Pp. 40-66.
- Wozniak, R. (1992). René Descartes y el legado del dualismo mente-cuerpo. René Descartes y el legado del dualismo mente-cuerpo.