



UNIVERSIDAD DE CHILE

UNIVERSIDAD DE CHILE  
FACULTAD DE FILOSOFÍA Y HUMANIDADES  
ESCUELA DE POSTGRADO

Las consecuencias de la distinción entre procesamiento computacional y cognición humana

Tesis para optar al grado de Magister en Filosofía

Ismael Frías Valderrábano

Profesor guía: Rodrigo González Fernández

Ciudad de México, año 2020

## Índice

<b>INTRODUCCIÓN .....</b>	<b>3</b>
<b>CAPÍTULO I: LA VISIÓN COMPUTACIONAL DE LA MENTE .....</b>	<b>7</b>
<b>EXPLICACIONES MECANICISTAS DE LA MENTE .....</b>	<b>7</b>
¿QUÉ ES COMPUTAR? .....	22
EL FUNCIONALISMO, UN PRINCIPIO ANTI BIOLÓGICO .....	32
<b>CAPÍTULO II: LOS LÍMITES DE LA COMPUTACIÓN FRENTE A LA MENTE .....</b>	<b>38</b>
INTELIGENCIA E INTENCIONALIDAD .....	38
EL ARGUMENTO DE LA HABITACIÓN CHINA DE SEARLE .....	48
LA INTENCIONALIDAD COMO CONDICIÓN NECESARIA PARA LA COMPRESIÓN LINGÜÍSTICA .....	56
<b>CAPÍTULO III: LA EXPLICACIÓN BIOLÓGICA DE LA MENTE Y DEL ENTENDIMIENTO HUMANO .....</b>	<b>61</b>
LA CONCIENCIA, UNA CARACTERÍSTICA PRIMARIA DE LA MENTE .....	61
LA APROXIMACIÓN COMPUTACIONAL DIGITAL COMO UNA EXPLICACIÓN INSUFICIENTE DE LA MENTE .....	69
LAS IMPLICACIONES ONTOLÓGICAS DEL FUNCIONALISMO Y SU ESCISIÓN DEL MUNDO .....	76
LA REALIZACIÓN BIOLÓGICA DE LA MENTE DENTRO DE UNA ONTOLOGÍA MONISTA .....	80
<b>CONCLUSIONES.....</b>	<b>86</b>
<b>BIBLIOGRAFÍA.....</b>	<b>92</b>

## Introducción

En esta tesis se pretende analizar la teoría computacional de la mente para discernir si explica de modo suficiente qué es la mente humana. Se exploran los principios que dan fundamento a la aproximación mecanicista de la mente desde sus comienzos con Charles Babbage y su pretensión de realizar trabajo mental por medio de la mecanización de sus procesos. Asimismo, se presentan los fundamentos de lo que hoy conocemos como computación digital gracias a Alan Turing y su famosa máquina teórica. Se abordan las disimilitudes entre la máquina diferencial de Babbage con respecto a la máquina de Turing, haciendo especial énfasis en que la primera es una máquina que, efectivamente, hace computaciones gracias a la disposición de sus engranajes y su diseño. Mientras que la segunda es una máquina teórica que no requiere ningún tipo de material para ser implementada. Su característica teórica radicaría en su naturaleza algorítmica, comprendiéndola como una serie de pasos finitos que se realizan recursivamente hasta llegar a un fin deseado. Además, se apunta que, gracias a las investigaciones de Augusta Ada, Condesa de Lovelace, la segunda máquina de Babbage, la máquina analítica, ella logra vislumbrar lo que hoy conoceríamos como una máquina multipropósito, ya que podría ser utilizada para realizar distintos tipos de operaciones matemáticas, pero también se podía usar para jugar juegos como el ajedrez o incluso para hacer música. La razón de ello es que esta máquina, de haber sido terminada, podría haber contado como la primera máquina digital, pero el énfasis que se debe de hacer es que Ada ya entiende que al asumir las capacidades mecánicas de este aparato podía enfocarse en sus capacidades matemáticas sin tener que estar pensando en el nivel de la implementación.

También se analiza la noción de computación de forma tal que su definición no pueda ser trivializada. Searle hace un reto a esta noción argumentando que bajo ciertas definiciones cualquier objeto podría caer bajo la categoría de computación. Ante ello, muchos filósofos, tratan de dar una definición no trivial para la computación. Para ello se toma la visión de Block y Copeland (Block, 1995a; Copeland, 1996a) porque ambos están conscientes del problema que Searle propone examinar. Se analiza qué es computar comenzando por establecer, como bien argumenta Copeland, que debe haber una relación de modelaje para que un objeto pueda ser considerado computacional, lo que se tiene es un criterio para

reconocer un sistema computacional en función de que éste no puede surgir sin un diseño. Esta relación de modelaje consistiría en la atribución de funciones a mecanismos, que pueden ser tanto físicos como teóricos, a estados internos de la máquina que podrían contar en su expresión mínima como lo que se conoce como procesadores primitivos.

El principal problema que se plantea en esta tesis es el de la intencionalidad y cómo es que ésta se puede dar en un sistema computacional que corre sobre la base de procesadores primitivos. Podemos comprender la intencionalidad como la capacidad que tenemos de representar, es decir, que un estado mental sea acerca de algo. Una noción crucial que se atiende durante todo este texto es el planteamiento funcionalista de que el cerebro es un motor sintáctico que conduce un motor semántico. Lo que se pretende explicar con esto es que, si comprendemos al cerebro como un sistema computacional, entonces está determinado por ciertas estructuras formales que son sintácticas que pueden procesar contenidos semánticos, esto gracias a una relación isomórfica que tendrían las estructuras sintácticas del sistema con el contenido representacional que estaría procesando. Esta noción es la que sostendría teorías como la del lenguaje del pensamiento de Fodor y la gramática universal de Chomsky (Chomsky, 1957; Fodor, 1979).

Uno de los planteamientos que afrontan con más fiereza esta aproximación computacional de la mente, y estrictamente la noción de inteligencia artificial fuerte que sostiene que un programa puede comprender historias, es el experimento mental de la Habitación China de Searle (1980). En él se propone imaginar a una persona angloparlante dentro de una habitación a la cual se le entregan preguntas en chino, esta persona tiene a la mano un manual de manipulación de símbolos chinos al inglés meramente por su forma. La persona usa este manual como un programa computacional que gracias a su estructura sintáctica le permite procesar los símbolos, sin embargo, no entiende ni una palabra de chino. De esta manera, una persona que comprenda chino podría pensar que la que está dentro de la habitación también lo hace. La tesis que se sostiene por medio de este argumento es un sistema computacional que solamente procesa sintácticamente no puede comprender el lenguaje y, por ende, tampoco historias. Asimismo, se plantean varias respuestas a este problema y los contraargumentos más importantes.

El principal problema es el de como un sistema sintáctico podría tener intencionalidad. Pero parecería que la mera forma sintáctica que procesa símbolos con un contenido intencional

otorgado por humanos es ciega a los contenidos semánticos. El principal problema al que apunta Searle es el de la naturaleza de los estados mentales. Se trata de dar un criterio para que un estado sea mental aludiendo a que todo estado mental para ser considerado como tal tiene que ser un estado consciente. La razón por la que los estados internos de una computadora no pueden contar como mentales es porque el contenido semántico que procesan sintácticamente está otorgado por observadores, esto es, por nosotros; en consecuencia, su intencionalidad es derivada de nuestra intencionalidad primitiva.

La conciencia tiene tres características fundamentales, porque es: interna, cualitativa y subjetiva. Es decir que todo estado mental consciente se presenta internamente en el sentido de que está dentro de nosotros y que está en relación con otros estados mentales que conforman una red. Es cualitativo porque tiene una forma de sentirse, todo estado mental consciente tiene una textura, por decirlo de alguna manera, y esto es a lo que muchos filósofos contemporáneos se refieren con la discusión de los *qualia*, que es el término utilizado para una cualidad singular de los aspectos fenoménicos de nuestras vidas conscientes (Peirce, 2020). Y, la característica más importante, los estados mentales conscientes son subjetivos. Esto quiere decir que no hay manera en que otra persona pueda tener acceso a mis estados mentales de la misma manera en que yo lo hago, y viceversa. Estas características hacen de la conciencia irreductible en términos de que, en tanto que son subjetivos, tienen una ontología de primera persona, es decir, existen para uno solamente.

Existen filósofos como Daniel Dennett que tratan de eliminar la conciencia de la ecuación afirmando que no es más que una ilusión y que no tiene un poder representacional que sea relevante para los procesamientos computacionales, es decir, que la conciencia sería algo así como un residuo del proceso (Dennett, 1992). Este tipo de posturas son consideradas eliminativistas y están comprometidas con el materialismo. Podemos comprender el materialismo como una postura ontológica en la que todo fenómeno, aunque no lo parezca, puede ser reducido a una sola substancia: la materia. Sin embargo, aceptar ciertas posturas materialistas sobre la mente nos obligaría a negar la conciencia. En contraposición al materialismo tenemos la visión dualista. El dualismo se puede comprender de dos maneras, el dualismo de sustancias y el de propiedades. El primero se refiere a que existen sustancias completamente diferentes que son excluyentes entre ellas. Un tipo de dualismo famoso es el de Descartes, cuando por medio de un criterio de claridad y distinción, afirma que si yo puedo

pensar una sustancia clara y distintamente separada de otra entonces puedo asumir que están separadas, de ahí su distinción entre *res cogitans* y *res extensa*. El dualismo de propiedades, en cambio, asume que hay dos tipos de propiedades que no pueden ser reducidas la una a la otra, por ejemplo, que hay propiedades físicas y propiedades mentales.

Searle acepta que la conciencia tiene ciertas características por el tipo de existencia de sus estados mentales, sin embargo, no está en contra de que se les estudie como como estados epistémicamente objetivos (Searle, 1999). La teoría computacional de la mente fallaría, porque trata de reducir los estados mentales, como portadores de contenido semántico, que es interno, cualitativo y subjetivo, a símbolos que tienen intencionalidad en virtud de su forma sintáctica.

El problema no necesariamente tiene que ver con toda la aproximación computacional de la mente, sino que el tipo de cómputo con el cual se pretenden abordar los estados mentales conscientes, que sería un lente digital. Y en ese sentido la aproximación computacional digital de la mente no sería suficiente para explicar dicho fenómeno.

Como alternativa al Funcionalismo y a la Teoría Computacional de la Mente, Searle tiene una postura que llama “realismo biológico”. Con éste toma distancia de la creencia de que toda la realidad es una construcción social, más bien entiende que hay un tipo de realidad cuya ontología es independiente de un observador. La existencia de los estados mentales no sería ajena a la naturaleza, los estados mentales conscientes son caracterizados por su intencionalidad, y ésta, en tanto que es parte de la naturaleza responde a la causalidad y, por supuesto, tiene poderes causales. La intencionalidad tiene ciertas condiciones de verdad, es decir, que para que haya una relación entre nuestras representaciones y sus referentes debe haber relaciones causales apropiadas que satisfagan las condiciones de verdad.

La mente es indisociable del cerebro de la misma manera que el agua lo es con la liquidez. Pero esto no es un problema de propiedades, más bien hay una estructura biológica que permite que haya mentes en un nivel de realización superior del cerebro de la misma manera que si el sistema digestivo digiera, el cerebro y el sistema nervioso piensan.

Finalmente, en esta tesis se alude a una descripción que, si no deja de ser computacional, sí deja de ser digital. Y la aproximación tendría que ser estrictamente biológica, por ello se retoma la postura de Gualtiero Piccinini (2010) quien postula la computación neuronal en la cual una aproximación biológica podría dar cuenta del contenido semántico tomando en

cuenta las características sensibles del sistema nervioso que permitirían que los estados mentales conscientes se nos presenten como internas, cualitativas y subjetivas.

## Capítulo I: La visión computacional de la mente

### **Explicaciones mecanicistas de la mente**

Una de las muchas explicaciones de la mente es la noción mecanicista, esta podría ser caracterizada por la idea de que podemos, mediante mecanismos ya sean análogos o digitales, comprender y reproducir características y poderes cognitivos que presenta la mente humana. Esta forma de pensar tiene sus orígenes a mediados del siglo XIX, durante el auge de la revolución industrial.

Charles Babbage, considerado por muchos el padre de la computación<sup>1</sup>, fue un pilar para el desarrollo de la ciencia computacional tal como la conocemos hoy en día, y no solo eso, sino que sus aportes permitieron poder contemplar posibles explicaciones del pensamiento gracias a las intuiciones que arrojan dos de sus proyectos más importantes: La Máquina Diferencial y La Máquina Analítica (Priestley, 2011). Proyectos que, a pesar de no haber sido concluidos a causa de circunstancias económicas y políticas, sus principios dan génesis a la explicación mecanicista de la mente tomando en cuenta nociones medulares que, a la postre, derivarían en problemas concernientes a la ciencia cognitiva, las matemáticas y a la filosofía de la ciencia cognitiva como el funcionalismo, la teoría del lenguaje del pensamiento, el problema de la sistematicidad, el álgebra, entre otros. Los alcances de estas posibles explicaciones y teorías que son inspiradas por la noción mecanicista de la mente que pone en movimiento el trabajo de Babbage no terminan en los linderos de la ciencia cognitiva o la ciencia computacional, sino que tienen repercusiones importantes en la manera en que concebimos nuestra propia existencia como seres humanos y la manera en que nos relacionamos con el mundo, se convierte en un problema que tiene alcances filosóficos en diversas disciplinas como la ontología, la epistemología, la ética, etc.

---

<sup>1</sup> Con esto me refiero a la noción de computación contemporánea, ya que desde mucho tiempo antes de Babbage la computación era hecha a mano por seres humanos.

Al estar inserto dentro del contexto histórico del siglo XIX, en plena revolución industrial, Babbage fue un producto del espíritu de su época y del ambiente intelectual. En una Inglaterra con un crecimiento tecnológico y económico como la Revolución Industrial venían nuevos retos por superar, uno de ellos fue la eficiencia de la realización del trabajo. Lo que se buscaba era tener una mayor producción en menor tiempo, que además fuera segura. Como lector de Adam Smith (Priestley, 2011), Babbage se inspiró en el concepto de la división del trabajo, según ésta, la mejor manera de producir algo de forma serial estaría en la división de distintos trabajos más especializados. Por ejemplo, un trabajador que se dedica a hacer broches debía de realizar distintos procesos para cada parte y ensamblarlos, al final del día podía tener una producción de 20 broches, con la aplicación de la división del trabajo se podían realizar al menos 240, asegurándose que cada parte estuviera hecha con la precisión que solo podía tener un trabajador especializado en la técnica para realizar ese trabajo en específico.

Babbage desarrolla su primera máquina, la Máquina Diferencial, que computaba e imprimía tablas de cálculo y de navegación con mucha precisión. El motivo por el cual se desarrolla esta máquina es que las tablas de cómputo se utilizaban para realizar cálculos precisos para tareas de índole económica o de navegación, que en esa época eran realizadas por trabajadores humanos llamados “computadores” y un error por su parte podría representar desde problemas financieros graves hasta costar vidas, por ello Babbage se compromete con la empresa de mecanizar los procesos matemáticos para evitar el error humano. La máquina de Babbage estaba basada en la metodología de Gaspard Riche de Prony, un ingeniero a quién el gobierno revolucionario francés le encomendó estandarizar los pesos y medidas al sistema decimal. De Prony aplica la noción de la división del trabajo a la computación realizando una técnica matemática conocida como el método de las diferencias. Este método permitía realizar funciones logarítmicas y trigonométricas que serían calculadas empleando solamente operaciones más simples como la suma y la resta.

A pesar de que anteriormente ya se habían construido máquinas para realizar cálculos matemáticos, como las de Pascal, Leibniz y Schickard, ninguna de ellas había logrado la exactitud que sí pudo alcanzar la Máquina Diferencial. La principal razón de esto es que la máquina de Babbage ejecutaba un algoritmo completo sin la necesidad de intervención humana en el proceso de cálculo, más allá de que el usuario ingresara inputs, jalara una que otra palanca y revisara los outputs. Gracias a su capacidad de realizar cálculos de manera



algorítmica, en dónde la totalidad del proceso matemático estuviera automatizado mecánicamente. La diferencia más importante que tenía la máquina diferencial con cualquier otra máquina de cálculo que se haya desarrollado antes es que no solamente era una calculadora, sino que tenía codificado todo un proceso matemático complejo entero en su engranaje.

Sin embargo, la importancia de su invención más allá de los progresos que podía significar para la economía, la astronomía o la navegación, tiene que ver con una empresa que Babbage tenía muy clara: La mecanización de la mente. La división del trabajo no solamente era aplicada ahora al trabajo físico, sino que esa noción se extendió al trabajo mental, primero gracias al trabajo de De Prony, y luego con la visión que tuvo Babbage de poder expresar ese trabajo algorítmicamente y mecanizarlo, dejando de lado el trabajo mental humano para realizar esa misma labor con más celeridad y exactitud.

Babbage no solo trabajó en la Máquina Diferencial, sino que su inventiva lo llevó a imaginar y desarrollar, aunque sin la posibilidad de implementar, otra máquina mucho más poderosa: La Máquina Analítica. Su desarrollo comienza cuando Babbage decide reexaminar el mecanismo de la Máquina Diferencial y se propuso evaluar fórmulas algebraicas mecánicamente, para ello implementó una serie de cambios en el mecanismo, sobre todo con respecto a los ejes que llevaban los engranes para darles más movilidad y poder realizar operaciones más complejas, además de implementar un sistema de lectura de tarjetas perforadas inspirado en los telares de Jacquard, mediante las cuales las variables eran ingresadas. Una característica importante de esta máquina tiene que ver con las motivaciones filosóficas de Babbage con respecto a la notación de las matemáticas en donde hace una distinción entre operaciones y variables, esto tuvo una repercusión importante ya que los inputs de la máquina eran variables ingresadas a la máquina y operaban distintas partes, por ello también podemos decir que la productividad se vio potenciada por la distinción de Babbage en tanto que no solamente podía realizar una sola operación sino una gran cantidad de procesos gracias a su capacidad combinatoria.

Al principio esta máquina estaba pensada solamente como una evaluadora de fórmulas algebraicas, sin embargo, sus implicaciones fueron mucho más importantes que esa función, en el sentido de que lo que se tenía era una suerte de traducción del lenguaje del análisis a la maquinaria, esto gracias a la distinción mencionada anteriormente ya que se podía ver en dos

niveles; el primero sería la estructura física de la máquina que era un reflejo del segundo nivel, la estructura del universo matemático. Las tarjetas eran una mera traducción de fórmulas algebraicas que al ingresarse en forma de variables controlaban la disposición de los estados internos de la máquina.

Ada Lovelace, quien mantiene una constante comunicación con Babbage durante el desarrollo de la Máquina Analítica, se da cuenta del potencial matemático de este proyecto gracias a la distinción de Babbage, lo que le permite establecer otra distinción mucho más importante entre lo matemático y lo mecánico. Ella se concentra más en el nivel matemático, que ya daba por sentada la posibilidad de mecanizar procesos sin la necesidad de explicar cómo sucedía esto en el nivel de los engranes, para describir la manera en que las leyes analíticas podían ser organizadas y combinadas para causar todas las derivaciones de ese tema sin tener que entender las capacidades del mecanismo que ya eran asumidas (Priestley, 2011). Su visión le permitió ver el poder de la máquina más allá de lo que vislumbró Babbage en el sentido de que la pudo comprender como una máquina multipropósito que no solo evaluaba fórmulas. La comprendía como “la encarnación de la ciencia de las operaciones, construida con particular referencia a los números abstractos como el tema de esas operaciones” (Lovelace, 1843), mientras que la Máquina Diferencial solamente efectuaba una sola operación por cada configuración de sus estados internos y tenía un grupo de operaciones muy limitado. Las posibilidades que encontró Lovelace no solo se limitaron a las matemáticas, sino que vio que la máquina tenía un potencial enorme en tanto que podía tener usos en la música o en los juegos como el ajedrez, es decir, que los mecanismos podían ir más allá de la representación de la aritmética y los números. Esto no quiere decir que, de hecho, Lovelace ni Babbage contemplaran la posibilidad de la emergencia de una mente o de la mecanización de una mente en su totalidad, lo que vieron fue que se podían plasmar ciertos procesos realizables por la mente en un mecanismo. Como dice Lovelace, “La Máquina Analítica no tiene pretensiones de originar nada. Ésta puede hacer lo que sea que nosotros sepamos ordenarle para realizarlo” (Lovelace, 1842), asimismo, Babbage estaba consciente del problema que implicaba el pensar la mente como una máquina y sostiene que es posible crear inteligencia mediante la construcción de máquinas.

La importancia de la Máquina Analítica consiste en que demuestra cómo ciertos procesos que se pensaba solamente pertenecían al campo de la cognición humana podían ser

mecanizables, entre ellas la capacidad de realizar operaciones matemáticas. Además, abre una pregunta que Babbage no dejó pasar por alto ¿Es pertinente utilizar ciertos términos antropomórficos para caracterizar las capacidades de la máquina tales como memorizar o saber algo?

Si podemos caracterizar de alguna manera la noción de la mente mecanizada, tenemos que mirar hacia sus inicios con Babbage. Su máquina marca un hito dentro de la historia de la ciencia de la computación, en el sentido más fundacional. Sus logros sientan las bases para comprender los inicios de las computadoras contemporáneas, pero, sobre todo, la capacidad de representar en términos mecanicistas un proceso de resolución de problemas matemáticos abre las puertas a un mundo de posibilidades, ya que parecería que con estos avances se podría aspirar a mecanizar la mente, y en ese sentido, también tener una mayor comprensión de su funcionamiento.

La Máquina Diferencial, como un primer intento de mecanizar una capacidad del intelecto, tiene codificado un proceso de cálculo matemático en sus engranajes, es una máquina más bien análoga, pero que sienta las bases para poder pensar que es posible reproducir ciertas capacidades mentales, y la reproducción implicaría una comprensión de las formas del proceso para poder ser plasmadas en un mecanismo. La Máquina Analítica, gracias a la distinción entre operaciones y variables, tiene una mayor flexibilidad en tanto que presenta un rasgo productivo, podríamos decir que es sistemática, a diferencia de la Máquina Diferencial, en el sentido de que no encarna una sola operación, sino que la configuración de sus estados internos, o sus engranajes, además de un innovador sistema de tarjetas perforadas, permite que la máquina se pueda organizar de distintas maneras tal que pueda realizar más de una operación, en este sentido, podríamos contemplar en ella una suerte de proto-programación.

Alan Turing, otro hito dentro de la historia de la computación, comenta con respecto a la objeción de Lovelace (Turing, 1950a), en el cual las máquinas están determinadas por aquello que nosotros les ordenemos hacer, que ella ni Babbage tuvieron en su época la evidencia como para alentarlos a pensar que una máquina podría, por ejemplo, aprender. Sostiene que la Máquina Analítica ya tenía las propiedades de una máquina universal y en el mismo sentido que Lovelace la pudo vislumbrar como una máquina multipropósito, él ya concibe

esta máquina como una máquina digital muy básica, pero que cuenta con los principios para serlo si tan solo tuviera una velocidad y una capacidad de almacenamiento más grandes.

La importancia de Turing dentro de la discusión sobre la noción computacional de la mente es enorme. Él contribuye a esta disciplina con la Máquina de Turing y el Test de Turing, y es por medio de la primera que se establecen los criterios de la computación digital dentro del marco en el cual se trató el *Entscheidungsproblem* de David Hilbert en 1930, el cual se proponía la posibilidad de determinar si una fórmula lógica, de primer orden, es universalmente lógica, es decir, que se pretendía determinar un procedimiento algorítmico que mediante una función computable pudiera decidir si los enunciados matemáticos son demostrables. La respuesta de Turing fue negativa, pero permitió comprender la concepción de computación mediante la Máquina de Turing. Ésta no es una máquina como las de Babbage, de hecho, es muy diferente. Se podría decir que las máquinas de Babbage eran análogas con respecto a la de Turing. La máquina de Turing puede ser implementada, pero no es su realización material lo que importa, es, más bien, una máquina teórica de carácter abstracto fundada en qué es un algoritmo. Mientras que las máquinas de Babbage encarnaban en sus engranajes la capacidad de realizar determinados cálculos matemáticos, la Máquina de Turing no requiere ser implementada en ningún material específico ni con alguna forma especial, lo único que importa es que pueda correr un algoritmo. Para Turing la inteligencia no se encuentra en la máquina, sino en un programa computacional, y justamente la Máquina de Turing es eso. Por ello Turing no discrepa mucho de los comentarios de Ada Lovelace con respecto a la capacidad de la Máquina Analítica de Babbage, ya que en ellos comenzamos a vislumbrar esta diferencia entre el programa y la máquina, que hace Turing en la distinción entre lo mecánico y lo matemático que propone Lovelace.

Una Máquina de Turing no necesariamente debe ser concebida como una máquina con engranajes y palancas, de hecho, se caracteriza por ser una máquina teórica, esto quiere decir que es la idealización de un proceso que computa símbolos de manera algorítmica.

Podemos comprender un algoritmo como un conjunto de reglas que cuando son aplicadas a un problema lo pueden resolver recursivamente mediante pasos finitos. Podemos encontrar algoritmos no solo en las matemáticas, sino que están en muchos procesos de resolución de problemas. Por ejemplo, tengo mucha sed y estoy en una plaza pública; el problema a resolver sería encontrar agua y las reglas serían “ve a buscar agua a una tienda, si hay, entonces

compra, si no hay, ve a la siguiente tienda hasta que la encuentres”, así puedo encontrar agua y solucionar mi problema mediante la implementación de manera recursiva hasta que tenga el resultado deseado. De la misma manera, uno podría resolver problemas mucho más abstractos, un ejemplo muy común es el del algoritmo de Euclides en donde se busca el máximo común denominador de dos números. Las reglas por seguir serían las siguientes:

- i) Dividir A entre B, anotar el resultado y el remanente
- ii) Si el remanente es igual a 0, entonces el resultado anotado es el resultado final y debemos parar.
- iii) Si el remanente no es igual a 0, entonces dividir B y el remanente para ejecutar de nuevo el paso i.

Al aplicar el algoritmo para encontrar el máximo común denominador de, por ejemplo, 99 y 27, obtendríamos lo siguiente:

<b>A</b>	<b>B</b>	<b>Resultado</b>	<b>Remanente</b>
99	27	3	18
27	18	1	9
18	9	2	0

Según las reglas que se dieron, el máximo común denominador de 99 y 27 sería 2.

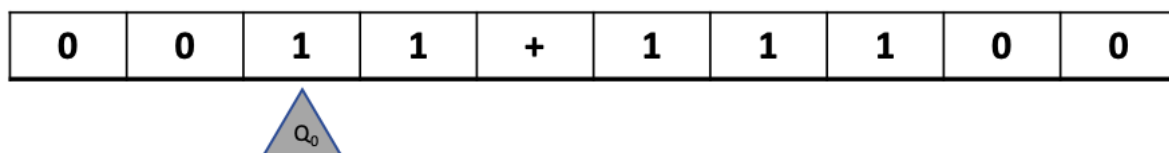
La Máquina de Turing define de manera abstracta y matemática qué es computar, y computa en términos de mecanismos. Si comprendemos el algoritmo como el procedimiento que se implementa mediante la Máquina de Turing, entonces esta es algorítmica. El algoritmo, cómo se vio anteriormente, tiene reglas y pasos finitos que se corren de manera recursiva. Esta recursión es necesaria para satisfacer una condición, y cuando ésta es satisfecha el algoritmo se detiene. Gracias a la capacidad de implementar algoritmos, la Máquina de Turing no solo se reduce a hacer una sola operación, sino que tiene la posibilidad de realizar tantas operaciones como podamos pensar de manera algorítmica, por ello se dice que la Máquina de Turing es una máquina universal, su potencia consiste en la posibilidad

de realizar tantos procesos como podamos darle ordenes, desde operaciones muy básicas hasta correr programas computacionales sumamente sofisticados.

Ahora que se ha expuesto cómo funciona un algoritmo y que se ha caracterizado a la Máquina de Turing como una máquina que corre procedimientos algorítmicos, podemos ver su funcionamiento. Una caracterización de la articulación de la máquina de Turing es la siguiente:

Usualmente, la Máquina de Turing se representa mediante una cinta infinita dividida en celdas, con una cabeza lecto-escritora, que se mueve hacia derecha e izquierda a través de la cinta, y que ‘recuerda’ alguno de los símbolos leídos o estados discretos (por contraposición a continuos). En un tiempo  $t$ , la cabeza que se encuentra en un estado interno ( $q_0, \dots, q_n$ ) lee el símbolo de la celda de la cinta ( $b_1, \dots, b_n$ ). En función del estado interno, de lo leído, y del microcódigo o programa de la cabeza, esta mantiene el símbolo o lo borra y escribe otro. Luego, se detiene o se mueve a otra celda, continuando las computaciones hasta que se detenga la cabeza, lo cual marca el final de procedimiento. (González, 2011a, p. 190)

Para poner un ejemplo, haremos una Máquina de Turing que realice la computación “2+3”. Entonces tendríamos una cinta de esta manera:



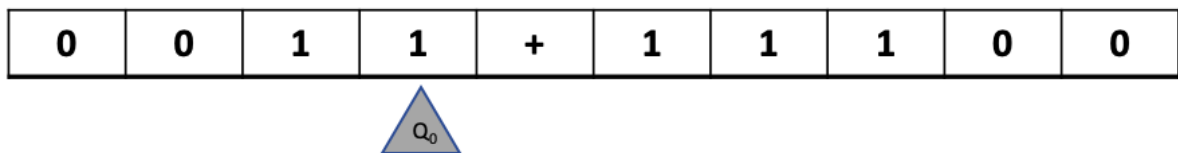
La cabeza lectora está determinada por el algoritmo, que en este caso es representado por una tabla, entonces tendríamos lo siguiente:

	$Q_0$	$Q_1$
1	1 D $Q_0$	0 Halt
+	1 D $Q_0$	
0	0 I $Q_1$	

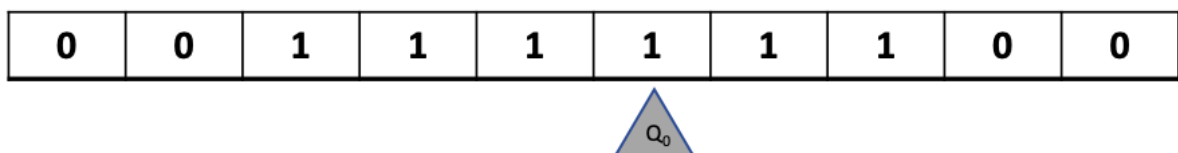
Esta tabla se leería de la siguiente manera:

- Si la cabeza está en el estado interno  $Q_0$  y lee el input “1”, entonces escribirá “1”, se moverá a la derecha (D) y permanecerá en el estado  $Q_0$ .
- Si la cabeza está en el estado interno  $Q_0$  y lee el input “+”, entonces escribirá “1”, se moverá a la derecha y permanecerá en el estado  $Q_0$ .
- Si la cabeza está en el estado interno  $Q_0$  y lee el input “0”, entonces escribirá “0”, se moverá a la izquierda (I) y cambiará al estado  $Q_1$ .
- Si la cabeza está en el estado interno  $Q_1$  y lee el input “1”, entonces escribirá “0” y se detendrá (Halt)

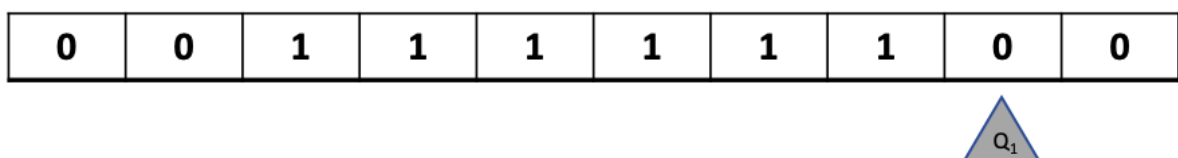
En la siguiente figura ya vemos como la cabeza leyó el dígito “1”, escribió “1”, se movió a la derecha y permaneció en el estado interno  $Q_0$ .



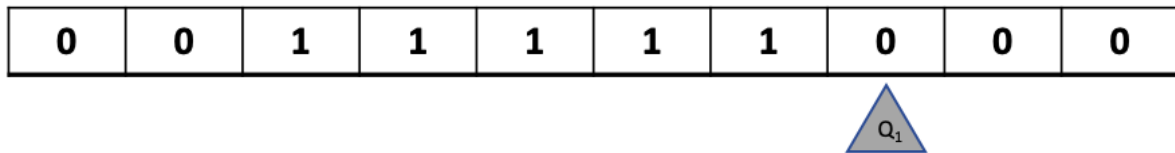
Cuando la cabeza se encuentre con el símbolo “+”, sucederá lo que previamente se había establecido en la tabla, se escribirá un “1”, se moverá a la derecha y permanecerá en el estado  $Q_0$  como vemos en la siguiente figura:



Lo mismo que en los primeros ejemplos, la cabeza se encontrará con un dígito “1”, sucederá lo que dicta el programa así que dejará escrito un “1”, se moverá a la derecha. Este movimiento lo hará recursivamente hasta que se encuentre con un dígito “0”, en este caso escribirá un “0”, esta vez se moverá hacia la izquierda y pasará al estado  $Q_1$ .



Una vez que la cabeza esté en el estado  $Q_1$  y lea el dígito “1”, entonces escribirá un “0” y parará. De esta manera, al final lo que tenemos es una sucesión de cinco “1” por lo que tendremos nuestra suma completa.



Como pudimos observar en el funcionamiento de la Máquina de Turing anterior, no se requiere de ningún tipo de engranaje que la haga funcionar, basta con tener lápiz y papel para poder implementarla, por lo mismo la máquina es meramente teórica e idealiza un proceso matemático; a la vez, podemos observar que no se requiere hacer uso de ningún tipo de facultad mental más allá de poder seguir ordenes, no se requiere ningún tipo de introspección o entendimiento para poder llevarla a cabo; la máquina está completamente determinada por las directivas que se le entregan, la cabeza lecto-escritora, que es la que realiza el proceso computacional, está totalmente determinada por el algoritmo que está implementando; el algoritmo siempre se da en forma de condicionales de tipo “si p, entonces q”, la máquina está completamente condicionada y, por lo mismo, no requiere ningún tipo de conciencia para funcionar.

Parecería que la Máquina de Turing es un concepto relativamente simple, pero tiene implicaciones importantes en el sentido de que escinde el funcionamiento de la máquina del material donde se implemente, en tanto que es una máquina idealizada lo que realmente importa es el programa o el algoritmo, por ello se le ha concebido como un antecedente del principio de realizabilidad múltiple, según el cual ser una máquina X es funcionar como esa máquina sin importar los materiales donde se implemente. Además, a pesar de su presunta simpleza, es un dispositivo extremadamente potente, sin él no viviríamos en el mundo que conocemos, ya que es la Máquina de Turing lo que hace que funcionen todas las computadoras que conocemos hoy en día, desde una laptop, un satélite, e incluso las llevamos en el bolsillo todos los días en nuestros smartphones. Su importancia consiste en que, a diferencia de la Máquina Diferencial de Babbage, este tipo de máquina tiene la capacidad de realizar cualquier proceso algorítmico siempre y cuando esté adecuadamente programada,



podríamos decir que en su funcionamiento tiene todos los procesos algorítmicos en potencia, mientras que la máquina de Babbage encarna en sus engranes un solo proceso. Como propone González:

El poder de una Máquina de Turing radica en que puede, en principio, simular cualquier sistema algorítmicamente calculable. Cualquier máquina que implemente un procedimiento efectivo M podrá ser imitada por una Máquina de Turing, y esto es supuestamente crucial para la construcción de máquinas programadas capaces de exhibir la misma inteligencia de un humano. (2011, p. 191)

Turing argumenta que, si una máquina X puede comportarse como otra máquina Y, entonces cuenta como tal. Por ello sostenía que una máquina podía pensar, si una máquina puede realizar ciertos procesos que hasta ese momento eran comprendidos solamente dentro del campo de la mente, como la inteligencia o aprender un lenguaje, entonces se podría decir que la máquina piensa de la misma manera que lo hace un cerebro, probablemente lo haga sobre cualquier material no biológico, pero en tanto que su algoritmo tiene la capacidad de mostrar procesos inteligentes, entonces es inteligente. Esta idea, que ya se venía gestando desde mucho antes que las máquinas de Babbage, sin embargo, fue la cereza sobre el pastel de la noción de la computadora digital como una metáfora de la mente la cual fue ampliamente trabajada después de los años 40. En esa década la idea de la información y de las computadoras electrónicas comenzaron a derivar disciplinas como la teoría informática, la inteligencia artificial y las redes neuronales artificiales, y prepararon el campo para la discusión interdisciplinaria y con ella el advenimiento de la “Revolución Cognitiva” (Frankish & Ramsey, 2012).

En los años 50, varios pioneros del ambiente investigativo de varias disciplinas ya estaban enfocados en lo que hoy comprendemos como arquitecturas cognitivas, en las cuales las representaciones y las reglas que especifican las operaciones en ellas consisten en símbolos discretos. La metáfora de la computadora digital como la mente abrió el camino para poder comprender algunos fenómenos de la psicología bajo las categorías del procesamiento informático, se desarrollaron mejores programas computacionales en la Inteligencia

Artificial, la gramática generativa en la lingüística, y la teoría del lenguaje del pensamiento en la filosofía<sup>2</sup>.

Dentro de este periodo destacan dos artículos presentados en el famoso Simposio de la Teoría de la Información del 11 de septiembre de 1956, que retoman la metáfora del computador y lo amplían, uno de Noam Chomsky y otro de Allen Newell y Herbert Simon. Ambos están de acuerdo con que se requiere una gramática generativa, la cual puede ser comprendida como una estructura que permita producir y comprender enunciados en cualquier idioma natural que, además, pueda dar cuenta de los procesos de adquisición y competencia del lenguaje, y sus raíces las podemos encontrar en la lógica simbólica como una continuación del trabajo de Frege, Russell y Whitehead a finales del siglo XIX.

En la lógica proposicional tenemos expresiones que están compuestas por proposiciones (generalmente representadas como “p”, “q”, etc.) que pueden ser V (verdaderas) o F (falsas) y conectivas lógicas como “ $\vee$ ” (disyunción), “ $\wedge$ ” (conjunción) y “ $\neg$ ” (negación), que pueden ser dispuestas de manera tal que podemos derivar reglas de inferencia como la condicional “ $p \rightarrow (q \vee p)$ ” que indica que si, ya sea, p o q son verdaderos, entonces p también lo es. O en la lógica de predicados, o lógica cuantificacional, las proposiciones son reemplazadas por predicados como “F” o “G” sobre uno o más argumentos, de los cuales uno es una constante o variable “x” o “y”, ambos son unidos por los cuantificadores “ $\forall$ ” (Para todo) y “ $\exists$ ” (al menos para un), la misma expresión que se manifestó anteriormente se podría leer en esta notación como “ $\forall x \forall y \forall xy \rightarrow \forall x \forall y (Gxy \vee Fxy)$ ” (Frankish & Ramsey, 2012).

Estas nociones de la lógica en la cual se tiene una sintaxis que le da forma a la manera en que las fórmulas comprenden las relaciones entre las proposiciones o los predicados nos permiten comprender cómo mediante implicaciones lógicas se podría procesar cierto tipo de información, que en este caso serían las condiciones de verdad de las proposiciones. El mismo Wittgenstein plantea las expresiones lógicas como operaciones, y éstas serían resueltas mediante la transitividad lógica de la misma manera que funciona una Máquina de Turing, que a pesar de correr los algoritmos mecánicamente, y en tanto que no importa el material en dónde se implementa, entonces la causalidad no es un factor importante, pero la transitividad sí.

---

<sup>2</sup> Este tema será abordado en el siguiente apartado con mayor detalle.

Chomsky tiene la idea de interpretar la gramática del lenguaje natural como equivalente a un mecanismo de la misma manera que la lógica proposicional nos otorga una sintaxis que permite operar símbolos, entonces ésta podría ser computable por una Máquina de Turing. Según su planteamiento se podrían generar oraciones de un lenguaje a partir de esas reglas. Él concluye que se necesita una gramática transformacional en la cual reglas de la estructura de frases como  $S \rightarrow NP$  y  $VP \rightarrow V(NP)$  generen la estructura profunda de la oración y luego las reglas transformacionales la alteren para obtener una estructura superficial. En ese sentido la gramática transformacional tendría el poder de una Máquina de Turing (Frankish & Ramsey, 2012). Gracias a esto, Chomsky ofrece una visión de cómo podrían verse las representaciones y las operaciones mentales montadas sobre una estructura sintáctica que permita manipular símbolos.

Asimismo, Newell y Simon concuerdan con Chomsky en la necesidad de tener una gramática generativa, sin embargo, ellos abordan el problema no desde la lingüística sino desde la computación. Ellos relacionan este tema con la Inteligencia Artificial por medio de las realizaciones físicas de los tipos de dispositivos explorados abstractamente en la noción mecanicista. Para entonces Claude Shannon ya había demostrado que los switches eléctricos podían ser dispuestos de tal manera que pudieran encenderse y apagarse ellos realizando así una operación matemática. No es que ellos hayan implementado esto por primera vez, de hecho, ya se tenía la primera computadora digital de propósito general, la ENIAC, en 1946 con la primera arquitectura serial Von Neumann, construida sobre la noción de la Máquina de Turing, la EDVAC. Sin embargo, ellos escriben el primer programa de Inteligencia Artificial, el Lógico Teórico, en el primer lenguaje de procesamiento de listas y lo corrieron en una computadora digital. La influencia de la lógica simbólica fue obvia en su tarea: descubrir pruebas de teoremas en lógica proposicional (Frankish & Ramsey, 2012).

Es en los años 50 que la ciencia cognitiva, adquiere su nombre gracias a la publicación de la revista Cognitive Science y de la Cognitive Science Society. Ésta aglomera varias disciplinas como la filosofía, lingüística, neurociencia, psicología, inteligencia artificial y antropología y trata de dar cuenta de la cognición por medio de explicaciones, éstas son posibles gracias a ciertos principios que se han adoptado para poder entender los procesos por medio de los cuales podemos conocer. Estos principios los tenemos gracias a la aplicación de categorías computacionales en el estudio de la cognición, antes de esa

aplicación la mente se presentaba como un fenómeno misterioso que en muchos casos era abordado mediante explicaciones ontológicas en dónde se hablaba de distintas naturalezas, como la distinción mente-cuerpo, en que dos entidades que funcionaban en niveles diferentes resultaban inconmensurables. Los avances de la informática permiten aplicar la metáfora de las computadoras como una mente a los procesos cognitivos comprendiéndolos como procesos computacionales, de esta manera la mente deja de verse como un misterio y se pretende dar cuenta de ella desde la ciencia.

El nombre de las explicaciones de los procesos cognitivos dentro de la ciencia cognitiva es “arquitecturas cognitivas” y las podemos comprender como propuestas generales sobre los procesos y representaciones que están operando dentro del pensamiento inteligente. Éstas se dedican a dar cuenta de aspectos importantes para la cognición humana como la resolución de problemas, memoria y aprendizaje (Frankish & Ramsey, 2012). Estos son abordados como programas computacionales por lo que no solamente sirven para explicar la cognición, sino que también han ayudado a disciplinas dentro de la ciencia cognitiva como la inteligencia artificial y otras ciencias computacionales para diseñar robots y computadoras que puedan poseer algunas habilidades cognitivas humanas.

¿Qué es una explicación en la ciencia cognitiva? De acuerdo con mucha investigación filosófica sobre la explicación, sostengo que las explicaciones científicas son generalmente descripciones de mecanismos que producen los fenómenos a explicar. Un mecanismo es un sistema de partes relacionadas cuyas interacciones producen cambios regulares. (Frankish & Ramsey, 2012)

Esta definición es importante en tanto que da forma constitutiva de una explicación en ciencia cognitiva, o arquitectura, en tanto que está conformada por partes, que dentro de las arquitecturas son entendidas como representaciones, que interactúan entre sí y que sus interacciones son los procesos computacionales. Las arquitecturas trabajan mecánicamente dependiendo de cómo es que las representaciones cumplen una función dentro del proceso. Podemos comprender una arquitectura como una propuesta sobre los tipos de representaciones mentales y los procesos computacionales que constituyen un mecanismo para explicar un rango amplio de tipos de pensamiento. Estas pueden ser entendidas de la

misma manera en que un programador escribiría un software que puede ser implementado en cualquier hardware.

En resumen, desde mediados del siglo XIX, gracias a la inventiva de Charles Babbage se pudo contemplar el trabajo mental como algo reductible a procesos que podían ser implementados en una máquina. Él trata de resolver el problema del error humano a la hora de hacer cálculos matemáticos y ahorrar tiempo en revisar las tablas de cálculo con su Máquina Diferencial, que encarnaba el método de las diferencias establecido anteriormente por De Prony. Esta máquina no puede ser pensada como una computadora digital, como nuestras computadoras contemporáneas, pero al menos pone sobre la mesa la posibilidad de realizar procesos que eran considerados propios de la mente en una máquina. Su Máquina Analítica ya presenta ciertas características que, como apunta Turing y Ada Lovelace, permiten realizar no solo una sola operación, sino que tiene la propiedad combinatoria que le permite poder ser programada por medio de tarjetas perforadas, esto gracias a la distinción que hace Babbage entre operaciones y variables. Lovelace se enfoca en su propia distinción entre lo matemático y lo mecánico dando por sentadas las capacidades de la máquina, lo cual le permitió enfocarse en las potencias matemáticas que se podían explorar gracias a la máquina, además de tener presente que la máquina no solamente podía ser utilizada para calcular, sino que también tenía usos recreativos, como el ajedrez o la música. Turing, quien ya tenía en mente su Máquina de Turing como una máquina idealizada, comprende el argumento de Lovelace, ya que la distinción entre lo matemático y lo mecánico concuerda con su idea de que lo que importa es el programa. A su vez, Turing, un siglo después, establece lo que se comprende por computar gracias a su Máquina de Turing, que mediante la implementación de un algoritmo puede realizar cualquier operación siempre y cuando esté bien programada, con ello nace la idea de la máquina universal que puede realizar cualquier operación, pero con ello también surge la idea de que una computadora puede emular las capacidades de la mente, y si una computadora realiza esas mismas funciones, entonces cuenta como una mente. Finalmente, una de las consecuencias de las ideas de Turing la computación tiene grandes avances, se desarrollan las primeras computadoras digitales eléctricas, y a la par la psicología y la lingüística comienzan a retomar ciertos principios computacionales para comprender la mente, como lo harían Chomsky y Newell y Simon. Cada quién en su respectiva disciplina, pero todos concordando en que es necesaria una

gramática generativa que pueda dar cuenta de los procesos psicológicos y lingüísticos que sucederían en nuestras mentes como parte de un proceso cognitivo, y estos son vistos como procesos lógicos que serían una continuación del trabajo de Frege, Russell y Whitehead. La ciencia cognitiva retoma por completo la metáfora de la mente como una computadora digital, sobre todo lo que comprendemos como las arquitecturas clásicas, en ellas se trata de explicar la mente mediante arquitecturas cognitivas que serían la manera en que un programador vería un programa de computadora, además de que tienen un valor explicativo en tanto que pueden explicar ciertos procesos mecánicamente. No obstante, como bien apunta Barbara Von Eckardt, la ciencia cognitiva aún no genera consenso y, en términos kuhnianos, aún está en una etapa pre-científica (Von Eckardt, 1996), sin embargo, sí hay una constante, que casi todas las arquitecturas han utilizado principios computacionales para caracterizar a la mente.

¿Qué es computar?

Una definición estándar de lo que comprendemos por computar es la que dicta que hacerlo es ejecutar un algoritmo, y más precisamente, decir que un dispositivo u órgano computa es lo mismo que decir que existe una relación de modelaje de cierto tipo entre él y una especificación formal de un algoritmo y una arquitectura que lo soporte. En ese sentido, la Máquina de Turing, en tanto que tiene una relación constitutiva con los algoritmos, también definiría lo que comprendemos por computar. Sin embargo, esta concepción se vuelve problemática ya que puede ser trivializada con cierta facilidad.

El filósofo John Searle tiene un argumento en el cual argumenta que una pared puede ser una computadora ejecutando el programa *Wordstar* en donde la pared está ejecutando una suma del tipo “pared + 1” donde “1” es el tiempo que transcurre, lo que él sostiene es que se puede describir un objeto de tal manera que bajo esa misma descripción el objeto puede ser una computadora digital, y que para cualquier programa y cualquier objeto lo suficientemente complejo hay alguna descripción de éste bajo la cual se esté implementando un programa. Otra versión de este mismo problema es el de la cubeta de agua de Ian Hinckfuss, en donde tenemos una cubeta transparente llena de agua en un día de primavera, a un nivel molecular están sucediendo muchas reacciones físicas, pero en un determinado momento todas las

reacciones comienzan a ejecutar un programa computacional humano. Lo que sostiene Hinckfuss es que cualquier objeto bajo ciertas condiciones físicas podría tener la suficiente actividad como para correr el programa de una mente, pero ¿esto realmente cuenta como un procesamiento computacional?

Parecería que el reto que nos ponen tanto Searle como Hinckfuss es el de dar una definición de computación que no pueda ser trivializada, sobre todo cuando lo que se trata de hacer en la ciencia cognitiva es dar cuenta de la cognición, que es una capacidad de la mente desde una perspectiva computacional. Según la noción de computación que se ha abordado a partir de Turing, computar tiene que ver con la posibilidad de realizar operaciones por medio de la manipulación de símbolos implementando un algoritmo hasta llegar a un resultado. Tanto Copeland como Gualtiero Piccinini abordan este problema (Copeland, 1996a; Piccinini, 2016a), la táctica que utilizan es la de abordar la definición de lo computacional sin comprometer su componente más distintivo, que sería que lo computable es aquello que puede ser procesado por una Máquina de Turing.

Establecer una definición de computar no trivial es de suma importancia si lo que se pretende es comprender la mente bajo los términos computacionales, tal como lo pretende hacer la ciencia cognitiva. Uno de los problemas más determinantes en la discusión de la ciencia cognitiva, si no es que es el más relevante, es el de la inteligencia, ya que parecería que ésta sería la característica más esencial de la mente.

La ciencia cognitiva asumió a la computadora digital como metáfora de la mente y la pretensión comprender la mente en términos mecanicistas. Sin duda, ese tipo de explicación puede resultar muy seductora gracias a la posibilidad de reducir los procesos cognitivos de tal manera que puedan ser comprendidos como operaciones mecánicas. Además, un factor histórico hace que la ciencia cognitiva surja con una postura en contra de la escuela de pensamiento psicológico del conductismo, la cual, en sus expresiones más radicales, niega los estados mentales, los cuales sí serían tomados en cuenta por los científicos cognitivos en tanto que los que se procesa computacionalmente son símbolos que pueden entrar como inputs, ser procesados mediante los estados internos de la máquina y el programa para, finalmente, obtener un output.

A mediados de siglo B.F. Skinner abogó por el condicionamiento operativo (en el cual reforzar un acto lo pone bajo control, por ejemplo, incrementar el rango con el cual una rata presiona una barra). [...] Skinner era conocido por su conductismo radical, el cual repudiaba las apelaciones a estados internos en explicaciones sobre el comportamiento. (Frankish & Ramsey, 2012).

La ciencia cognitiva defiende que una explicación conductista, que para los años cincuenta dominaba el campo investigativo de la psicología, no es suficiente para explicar la conducta inteligente, porque la explicación se queda en el nivel de las conductas sin tratar de explicar por qué suceden de la manera en que se manifiestan en términos de estados mentales que obedecen a causalidad. A propósito de la confrontación que hace la ciencia cognitiva al conductismo, Ned Block hace una distinción entre dos tipos de definiciones, por una parte define las clases estructurales como definiciones que podemos encontrar en el consenso lingüístico y refieren al significado de la palabra, por ejemplo, podemos definir el agua como un líquido inodoro, insabor e incoloro, son definiciones de palabras que podemos encontrar en los usos comunes de cualquier lenguaje y que podemos elaborar desde la comodidad de un sillón consultando nuestras intuiciones lingüísticas; en contraposición a este tipo de definiciones existe otra clase de definiciones, las funcionales, que son el resultado de una investigación, estas hacen referencia a su objeto en función de su estructura físico-química, que en el caso del agua sería H<sub>2</sub>O (Block, 1995a). Esta distinción sirve para marcar una diferencia esencial con respecto a la pretensión de reconocer la inteligencia según Turing, ya que el matemático realiza una definición de inteligencia en términos conductistas en su afamado artículo “Computing Machinery and Intelligence” (Turing, 1950). La ciencia cognitiva no puede aceptar esta definición porque su mirada está en aquello que causa las conductas causalmente, es decir los estados internos. Además de que una definición como la que nos otorga Turing no estaría basada en la experimentación requerida para satisfacer las condiciones empíricas sobre los estados internos que sí exige la ciencia cognitiva. En este sentido la definición de Turing es mucho más cercana a las clases estructurales que a las funcionales.

Block argumenta que Alan Turing, además de desarrollar la noción de la Máquina de Turing, plantea una definición de inteligencia con un criterio conductual mediante el Test de Turing. Existen distintas versiones de éste y su origen comienza con el Juego de la Imitación,



la primera versión consta de dos personas en habitaciones distintas, en una un hombre y en la otra una mujer, afuera debe haber un juez cuyo sexo no es relevante, el juez hará preguntas para determinar el sexo de las personas en las habitaciones, la mujer tendrá que responder con veracidad y la meta del hombre será engañar al juez respondiendo las preguntas como si fuera una mujer. Si el hombre logra engañar al juez, entonces el hombre habrá ganado el juego. En la segunda versión hay una computadora en una habitación, en la otra hay una mujer, y afuera está el juez; la mujer deberá dar respuestas verdaderas y la computadora contestará como si fuera una mujer para engañar al juez. Si el computador responde eficientemente a las preguntas que se le formulan, el resultado más probable es que el juez se equivoque y concluya que hay dos mujeres. Finalmente, la tercera es la versión estándar de lo que conocemos como el Test de Turing, en ésta hay una persona cuyo sexo es irrelevante en una habitación que responderá verazmente, y en la otra habitación hay una computadora que tendrá que engañar al juez haciéndose pasar por un ser humano. El argumento que subyace al Test de Turing tiene que ver con la creencia que él sostenía, la cual dicta que una computadora que pueda realizar las mismas operaciones que otra computadora, cuenta como tal. Por esto, si una computadora logra convencer al juez de que es un ser humano, entonces estaría presentando las conductas propias de un ser humano inteligente y, por lo tanto, presentaría inteligencia.

Si alguna máquina en particular puede ser descrita como un cerebro, nosotros solo tenemos que programar nuestra computadora digital para imitarlo y ésta también será un cerebro. Si se acepta que los cerebros reales, como los encontramos en animales y particularmente en los hombres, son un tipo de máquina, se seguiría que nuestra computadora digital adecuadamente programada, se comportará como un cerebro.  
(Turing, 1951)

González hace una crítica a Block porque caracteriza el Test de Turing como una definición de inteligencia, en la cual ser inteligente sería pasar el Test de Turing. En particular, González sostiene que en primera instancia, Turing no trata de dar una definición para evitar problemas metafísicos como el tener que delimitar qué es lo que se comprende por “inteligencia”, aunque también desestima el Test de Turing (al igual que Block) como un experimento científico válido basándose en la noción de experimento de Fodor.

Dada la concepción fodoriana de los experimentos existe un estrecho vínculo entre la verdad de una creencia y la posibilidad de que ésta sea producida causalmente, ya que un experimento no causa la creencia de que  $p$  a menos de que  $p$  sea el caso. Es por ello que los científicos usan los experimentos como artefactos cognitivos capaces de convencer al resto de la comunidad científica, o al club, de la solidez de una hipótesis. ¿Se da esta situación en el Test de Turing, esto es, puede contar como un experimento científico donde alguien es causado a creer que  $p$  si y solo si  $p$  es verdadero? (González, 2011a)

La respuesta de González es negativa, ya que el Test de Turing no nos induce a que  $p$  sea verdadero, sino todo lo contrario, nos llevaría a una falsedad en tanto que el computador simularía ser una persona. Este experimento convence a los jueces por medio de la simulación y por eso no satisface el requisito de causar creencias verdaderas. Sin embargo, si hay algo que González le concede a Block, es que a pesar de que Turing no aporta una definición, sí encuentra un componente conductista en el método propuesto.

Block argumenta que con el Test de Turing se define la inteligencia sin dar cuenta de las estructuras internas de aquello que podría ser considerado un proceso inteligente, ya que éste no puede contar como una investigación científica del tipo que nos llevaría a pensar que el agua es  $H_2O$ , es decir, no se explica en función de los componentes y las interacciones entre ellos. Argumenta que hay dos tipos de clases: Las estructurales y las funcionales. Las primeras serían clases naturales como agua o tigre, que tienen una esencia composicional oculta; y las segundas no tienen esencia con respecto a su composición, sino que son funcionales, porque su significado está dado por el rol que cumplen dentro de una estructura, por ejemplo, trampa para ratón o gen. Block argumenta que la inteligencia debe ser considerada como una clase funcional que debe ser investigada empíricamente, lo cual iría en contra de una definición conductual, consensuada. Lo que realmente importaría son los procesos y la manera en que las partes del fenómeno se relacionan entre sí para poder realizar el objeto que se estudia, por ello lo que debe ser determinado son las funciones involucradas en las clases funcionales, en este caso, la inteligencia.

Dada la importancia de la empresa investigativa sobre la inteligencia en la ciencia cognitiva bajo categorías computacionales, es necesario establecer el tipo de relación que tiene el modelaje y la especificación formal del algoritmo para no caer en una definición que

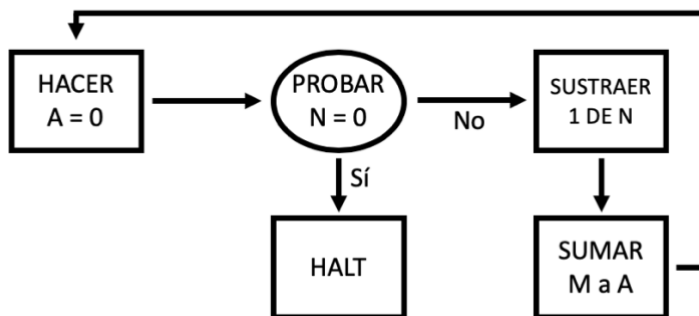
sea trivial, y que, por supuesto, no caiga en el territorio del conductismo. Por ello para poder caracterizar la computación tenemos que describir el procesamiento computacional.

El método utilizado para examinar la propiedad de la inteligencia es el análisis funcional, que fue desarrollado principalmente por Fodor y Cummins (Cummins, 1975; J. Fodor, 1965). Éste consiste en la descomposición recursiva de los procesos inteligentes en operaciones cada vez más básicas de tal manera que podamos contemplar el todo en virtud de sus partes. Un ejemplo de Block es el de los homúnculos:

Consideremos que la mente humana está representada en nuestra cabeza por un ser inteligente, un 'homúnculo'. Piense que este homúnculo está compuesto por homúnculos más pequeños y tontos, y cada uno de éstos están compuestos por homúnculos cada vez más pequeños y estúpidos hasta alcanzar un nivel de homúnculos completamente mecánicos (Block, 1995a)

Este método se plantea cómo una forma de analizar la manera en que funcionaría la inteligencia de la misma manera en que lo hace un computador, se apela a las competencias cognitivas más simples para explicar funciones cada vez más complejas y sofisticadas. De hecho, el análisis funcional es completamente compatible con la Máquina de Turing. Como se dijo en el primer apartado, la Máquina de Turing realiza operaciones mediante un procedimiento algorítmico y no requiere ningún tipo de introspección ni conciencia para funcionar, se ejecutan una serie de directivas de una manera mecánica. Al apelar a funciones cada vez más simples, la explicación de una operación compleja puede ser comprendida algorítmicamente gracias a microprogramas que son ejecutados estúpidamente por una computadora. Asimismo, Block pone un ejemplo que ilustra la manera en que una función puede ser desglosada en procesos cada vez más simples, asumamos que queremos multiplicar  $m$  por  $n$  como se muestra en la siguiente figura:

$$\begin{array}{ccc} \boxed{M} & \boxed{N} & \boxed{A} \\ m \times n & = & a \end{array}$$



El proceso de multiplicación sería el siguiente: El registro A está destinado al resultado de la operación, simbolizado por “a”. Una representación de 0 es colocada en el registro A, después en el registro N se prueba si contiene una representación de 0. Si la respuesta es sí, entonces el programa hace un Halt (alto) y la respuesta correcta es 0, pero si la respuesta es no, entonces el programa sustrae 1 de N de tal manera que el registro N ahora contiene una representación de n-1, y una representación de m es sumada al registro del resultado, A. Después el procedimiento hace un bucle hacía atrás hasta el segundo paso en donde se comprueba si de nuevo el registro N hay una representación de 0, si no, entonces de nuevo se resta 1, y de nuevo m se agrega al registro del resultado (Block, 1995a). Este procedimiento continúa hasta que N tiene el valor de 0, en ese momento m será agregado al registro del resultado exactamente n veces. En este punto, el registro del resultado contendrá una representación del resultado de toda la operación. Si representamos el procedimiento mediante una tabla, por ejemplo, haciendo la multiplicación de 7x5, entonces tendríamos lo siguiente:

M	N	A
7	5	0
	4	7
	3	14
	2	21
	1	28

	0	28
--	---	----

El cálculo es realizado mediante directivas de operaciones más simples como la suma, la resta y el Halt aplicadas recursivamente a modo de algoritmo, proceso que una Máquina de Turing podría tener.

Una implicación de comprometernos con la noción de que podemos comprender la mente como análoga a una computadora, y en específico la ciencia cognitiva que trata la propiedad de la inteligencia como un proceso computacional, es que las capacidades inteligentes pueden ser comprendidas a través de la descomposición recursiva de las operaciones cognitivas, lo que nos llevaría a un entramado muy complejo de capacidades cada vez menos inteligentes que están basadas, en última instancia, en procesos totalmente mecánicos y tontos.

Hasta ahora solamente se ha descrito cómo funcionan las operaciones en términos ideales, de la manera en que un programador puede contemplar un software, en virtud de un algoritmo que puede correr en una Máquina de Turing mecánicamente. ¿Pero qué pasa cuando tratamos de implementar materialmente nuestro programa? La respuesta está en los procesadores primitivos.

Un procesador primitivo es la parte más básica y primitiva de la máquina. Estos permiten que los programas puedan correr, son los mecanismos en donde pueden ser comprendidos los procesos algorítmicos. Para comprenderlos tenemos que ver cómo funcionan y hay que hacer una distinción importante entre lo simbolizado y el símbolo, por ejemplo, en la notación binaria, que es el símbolo que refiere a lo simbolizado, tenemos los numerales “1” y “0”, que en notación decimal también cuentan como 1 y 0, pero cuando se simbolizan otros números, como el 2, entonces tenemos “10”. Generalmente se utiliza la notación binaria para realizar una operación en una Máquina de Turing. Tenemos el ejemplo de la sumadora, la cual resuelve cuatro problemas:

$$0 + 0 = 0$$

$$1 + 0 = 1$$

$$0 + 1 = 1$$

$$1 + 1 = 10$$

Las tres primeras operaciones son verdaderas tanto en binario como en decimal, pero la última *solo* es verdadera cuando la expresamos en notación binaria. Cuando nosotros ingresamos un símbolo a modo de input a una computadora lo hacemos mediante una notación que la computadora interpreta, realiza la operación y arroja un output, que en el caso de una operación matemática sería el resultado. Sin embargo, tenemos que analizar qué es lo que hace la computadora cuando interpreta los símbolos a un nivel material. La Máquina de Turing en tanto que es una máquina teórica nos daría los procesos que nosotros podemos implementar en cualquier material, no es relevante que sea una tira de papel o conejos en cajas, lo único que importa es que uno pueda ingresar inputs que cuenten como símbolos a procesar. El mecanismo que se utiliza es el de las compuertas, por ejemplo, una compuerta Y es un dispositivo que acepta dos inputs y emite un solo output. En una compuerta excluyente, que puede ser comprendida como la conectiva lógica “V” (o), uno ingresaría dos inputs, si los dos son “1”, entonces la compuerta arroja un output “0”, pero si ingresamos un “1” y un “0” sin importar el orden, entonces la compuerta nos dará un “1”. Lo que hace este tipo de compuerta es detectar las diferencias, emite un “0” si no hay tales, pero si los inputs son distintos, entonces emite un “1”. Otro tipo de compuerta sería la inclusiva, en la cual la compuerta también detectaría las desigualdades, entonces si ingresamos dos inputs distintos, el output será un “0”, de igual manera que si ingresamos dos inputs “0//”, pero si ingresamos los inputs “1/1”, entonces el output de la compuerta será “1”. Todo lo que importa de una compuerta desde una perspectiva computacional es su función input-output (Block, 1995a).

El lugar en donde se encuentra el software con el hardware es en la manera en que ingresamos los inputs, pues las compuertas trabajan con estados físicos, por ejemplo, podemos implementar una Máquina de Turing en circuitos eléctricos y existen estados físicos que cuentan como representaciones dentro del proceso computacional. Así, podemos decir que 4 voltios cuentan como un “0” y 7 voltios cuentan como un “1”. La materia del hardware es completamente irrelevante en una descripción computacional, podríamos implementar el mismo algoritmo con compuertas donde haya gatos y ratones o con viento pasando por habitaciones a distintas velocidades, pero serían equivalentes computacionalmente, lo que importa son las funciones, y este es el fundamento de un principio computacional llamado funcionalismo.

Los procesadores primitivos son la parte más básica de una operación computacional, pero de la misma manera en que se vio en los ejemplos de la sumadora y de la multiplicadora, uno puede valerse de las funciones más simples y hacerlas interactuar de manera tal que obtengamos funciones más complejas, es decir, que los procesadores primitivos componen procesadores complejos, y estos pueden seguirse combinando para hacer procesadores cada vez más refinados. La frontera entre los niveles de procesadores puede ser porosa y es complicado establecer un límite entre ellos, de la misma manera en que no sabemos perfectamente en dónde comienza y termina la física y comienza la química.

Podríamos decir que una explicación mecanicista es aquella que da cuenta de un fenómeno en función de las partes que lo componen sistemáticamente, de forma tal que al establecer las relaciones entre las partes se pueda explicar el fenómeno en su totalidad. Las operaciones primitivas cambian de una arquitectura a otra, por ejemplo, podemos tener la arquitectura de la máquina de Turing, pero a la vez podemos tener una arquitectura de redes neurales con un patrón de conectividad y ciertos pesos en sus conexiones, en este caso los pasos son llevados en paralelo. Aquí cabe resaltar que la ejecución de un algoritmo no es necesariamente un procedimiento secuencial.

Existen algoritmos que son específicos para ciertas arquitecturas, esto no solo quiere decir que la máquina con esa arquitectura puede correrlo, sino que también cada instrucción que se da evoca explícitamente a la realización de alguna secuencia de operaciones primitivas hechas en esa arquitectura específica. Un algoritmo que al menos en una de sus instrucciones pida explícitamente hacer una multiplicación no puede ser específico para una arquitectura que solo tenga adición, pero no la multiplicación como un proceso primitivo, sin embargo, un programa que pida multiplicaciones puede correr esa arquitectura solo porque la instrucción de la multiplicación puede ser reemplazada por una serie de adiciones.

Copeland argumenta que para tener una definición de computación que no pueda ser trivializada debe de haber una relación de modelaje entre la máquina y su arquitectura (Copeland, 1996a). Es decir que debemos tener la capacidad de establecer una relación de identidad entre una arquitectura, o una descripción de la máquina, y una entidad  $\alpha$ , por ello Copeland apunta que la respuesta está en efectuar un sistema de etiquetado para  $\alpha$ .

El sistema consiste en dos partes: La primera es la designación de ciertas partes de la entidad como portadores de etiquetas, y la segunda es el método para especificar la etiqueta

llevada por cada parte del portador en cualquier momento determinado. Las etiquetas constituyen un código tal que secuencias espaciales o temporales de etiquetas tengan una interpretación semántica. En la explicación del comportamiento y la función de la entidad etiquetada uno adscribe la interpretación semántica asociada con las etiquetas directamente con los estados de etiquetado mismos. Es decir que el sistema de etiquetado es lo que permite asociar los procesos de  $\alpha$  con funciones que se entretajan y cumplen las órdenes que se les da haciendo, esto nos permitiría contemplar todas las partes como un programa computacional.

La noción de computación definitivamente tiene que ver con la manipulación de símbolos de manera algorítmica hasta lograr un resultado deseado. Y algo que me parece importante remarcar aquí es que uno puede tomar ciertos estados físicos de un objeto y asignarles un dígito, estos pueden ser latidos, pulsos, ritmos, etc. Pero hay que hacer hincapié en el hecho de que la asignación resulta arbitraria, es decir que la forma de la asignación depende enteramente de nosotros y nuestros deseos. En el caso de una computadora también hay una asignación de valores que responde a un interés específico, por ejemplo, hacer una suma o una multiplicación. Es decir que para establecer una relación de modelaje entre una arquitectura o la descripción de una máquina sin duda tenemos que poner atención en el sistema de etiquetado, la razón de esto es que ese sistema funciona como un puente que otorga la identidad de nuestra computadora como tal en función de los intereses con los que haya sido construida. Por eso no podemos decir que el universo sea un computador que se ejecuta a sí mismo al ejecutar las leyes de la física.

### El funcionalismo, un principio anti biológico

La noción computacional de la mente se sostiene bajo la premisa de que la mente puede ser comprendida bajo el esquema de un programa computacional. Como se vio anteriormente, la computación –comprendida como la manipulación de símbolos mediante un procedimiento algorítmico en el cual se ingresa un input y se obtiene un resultado en forma de input y que, además, responde a una relación de modelaje entre la máquina y su descripción, que a la vez refleja el interés por resolver un problema– tiene la bondad de que permite comprender la mente bajo su lente, esto quiere decir que si uno trata de comprender la mente bajo este



esquema, entonces nos damos cuenta de que la mente puede ser reducida a procesos mecánicos.

En el apartado anterior se habló de que para no tener una definición trivial de la computación es de vital importancia tomar en cuenta la relación de modelaje entre la máquina y su descripción, ya que de otra manera cualquier objeto podría ser una computadora porque prácticamente se puede definir cualquier objeto como tal. Esa relación de modelaje es sumamente importante no solamente porque establece una relación de identidad entre un objeto y su modelaje de tal manera que pueda ser comprendido como una computadora. Sin embargo, también es importante apuntar a algo: La relación de modelaje se funda en el funcionalismo.

Anteriormente se habló del funcionalismo como un principio de la computación, ya que éste dicta que uno puede descomponer un proceso de manera recursiva hasta llegar a los procesadores primitivos de la misma manera que ilustra Block en su ejemplo de los homúnculos (1995). Otro buen ejemplo que se usa en ese mismo artículo es el de cómo entendemos el lenguaje:

Parte del sistema reconocerá las palabras individuales. Este reconocedor de palabras podría estar compuesto de tres componentes: uno que tenga la tarea de capturar cada palabra entrante, una a la vez, para luego pasarlas al segundo componente. Este segundo componente compara la palabra objetivo con las palabras en su vocabulario (quizás ejecutando comparaciones simultáneamente). Cuando se encuentra una coincidencia envía una señal a un tercer componente que recupera la información sintáctica y semántica guardada en su diccionario. (Block, 1995)

Aquí el análisis funcional se aplica apelando a competencias cognitivas más simples para explicar una función más compleja. La idea de este tipo de explicación proviene de la manera en la cual funciona una computadora, es decir, por procedimientos algorítmicos, de la misma manera en que se ve en el ejemplo de la multiplicadora.

La manera de establecer una relación de modelaje entre una máquina y su descripción, como se vio en el apartado anterior, se logra a partir de un sistema de etiquetado. En este sistema lo que se está haciendo es establecer las funciones de cada uno de los elementos que constituyen el aparato en su totalidad. Por ejemplo, cuando se construye una computadora y

decimos que 4 voltios cuentan como un 0, y 7 voltios cuentan como un 1, lo que se está haciendo es atribuir un valor simbólico a un estado físico, y en ese sentido, muy a la Wittgenstein, el significado que se le atribuye a los estados físicos es el de su uso, que para nuestros fines es una función. A los voltios que corren por un circuito se les asigna una función que es la de un símbolo, ya sea 1 o 0, dentro de un sistema. Y lo mismo sucede con un circuito, por ejemplo, lo que conocemos por una compuerta Y no deja de ser un estado físico, que es un circuito con determinada forma, al cual se le atribuye la función de compuerta y aquél que establece la relación de modelaje es quien determina su uso, ya sea una compuerta inclusiva o exclusiva.

Una vez que se establece el sistema de etiquetado en los procesadores primitivos es posible construir funciones complejas, y en ese mismo sentido cuando se dota de funciones a los procesadores primitivos y a los estados físicos que están computando en forma que símbolos es posible construir símbolos con más símbolos. Cuando Block expone la posibilidad de comprender la mente como el software del cerebro también habla de que es posible comprender el cerebro como un motor sintáctico que lleva un motor semántico. De la misma manera que Copeland establece una distinción entre número y numeral (Copeland, 1996), Block hace la misma distinción entre número y símbolo, por ejemplo, podemos tener el número 1 y, por su parte, el símbolo "1". Una vez que se establecen las funciones en un sistema de cómputo los estados físicos son dotados de una semántica en tanto que son interpretados como símbolos, que, en el caso de una calculadora, refieren a números. Los inputs son considerados como si se refirieran a números. Pero una vez que nos introducimos, por ejemplo, dentro de la sumadora, debemos considerar que los estados binarios se están refiriendo a símbolos en sí mismos. Podríamos decir que el estado de cosas dentro del hardware está teniendo un contenido semántico en tanto que estos estados están funcionando como símbolos.

Si descomponemos funcionalmente un algoritmo hasta llegar a los procesadores primitivos, podemos notar que hay un cambio en el contenido: Los inputs y outputs que ingresamos los seres humanos refieren a números, pero los inputs y outputs de las compuertas refieren a numerales. La operación de componentes a un nivel superior funciona como un programa o algoritmo que manipula números, pero los niveles inferiores, como en los procesadores primitivos, no pueden ser explicados en términos numéricos, pero sí simbólicos.

Para ver nuestra calculadora como un motor sintáctico que lleva un motor semántico se requieren dos funciones: La primera es que haya un mapeo de números sobre otros números, y la segunda es que se haga un mapeo de símbolos sobre otros símbolos, sin fijarse en aquello a lo que refieren.

“La idea es que se interprete algo físico en la máquina o sus outputs como si fueran símbolos, y algún otro aspecto físico de la máquina como si indicara que los símbolos son inputs u outputs” (Block, 1995)

En este sentido, argumenta Block, que lo que tenemos son dos funciones: La primera es una función semántica, que tiene que ver con aquello a lo que refieren los símbolos o, en palabras de Copeland, a los números; y la segunda es una función sintáctica, que se refiere a los símbolos por sí mismos, o los numerales. Block sostiene que hay una relación isomórfica entre estas dos funciones porque hay aspectos físicos que pueden ser interpretados simbólicamente. Un dispositivo que manipula símbolos puede sumar números gracias a que “hay regularidades simbólicas isomórficas respecto a las relaciones racionales entre los valores semánticos de los símbolos”.

La posibilidad de establecer este tipo de relaciones mediante un sistema de etiquetado estaría íntimamente ligado al principio funcionalista. El análisis funcional refiere a un tipo de explicación, y derivadamente a un método investigativo que busca explicaciones de tipo funcionales. La explicación funcional es aquella que se sostiene en la descomposición recursiva de un sistema en sus partes componentes, esta explica el funcionamiento del sistema en los términos de las capacidades de las partes y la manera en que éstas se integran entre ellas (Cummins, 1975).

Ahora bien, dentro de la ciencia cognitiva el funcionalismo computacional, el término funcionalismo se aplica en un caso especial de explicación funcional. Lo que se trata es de definir una noción que viene de la psicología desde un modelo computacional, cualquier misterio de nuestra vida mental podría, al parecer, ser disuelto por medio del análisis funcional de los procesos mentales al punto de que puedan ser vistos como computaciones que en sus niveles mecánicos más bajos puedan comprenderse como los procesadores primitivos de una computadora. La noción clave de este tipo de funcionalismo es que se avoca a definir qué es una representación y un cómputo (Block, 1980).

La psicología, dentro de la ciencia cognitiva, sería contraria a la escuela del conductismo y pretende dar cuenta de aquello que causa las conductas, esto serían los estados mentales a modo de representaciones. Los estados mentales, o psicológicos, se entienden como representaciones del mundo a través del lenguaje del pensamiento, y los procesos psicológicos son vistos como cálculos que implican a estas representaciones.

Block explica que a pesar de que una computadora tiene contenido semántico, la máquina como tal no tiene intencionalidad, la computadora solamente está procesando símbolos como los numerales, pero no tiene idea de a qué es a lo que se refieren. “Aunque el modelo computacional de la mente tiene una explicación natural y directa de la inteligencia, no existe ninguna explicación para la intencionalidad que venga naturalmente con ella” (Block, 1995). Podemos entender por contenidos intencionales a los significados de nuestras representaciones internas. Sin embargo, bajo esa noción no existe un acuerdo sobre si el tema de si nuestro lenguaje interno es el mismo o es diferente del lenguaje con el que hablamos. Tampoco lo hay con respecto a la dirección en que un lenguaje se reduzca al otro (ya reducción del lenguaje natural al lenguaje del pensamiento o viceversa), es decir cuál es más básico, el contenido mental o el significado de sus símbolos internos.

Para Fodor, uno de los más importantes filósofos que postulan el lenguaje del pensamiento, argumenta que el significado del lenguaje se deriva del contenido del pensamiento, y que, por su parte, el contenido del pensamiento se deriva del significado de los elementos del lenguaje del pensamiento. Esto quiere decir que un estado intencional, como lo sería creer o desear algo, es un estado que responde a estar en una u otra relación computacional con otras representaciones (Fodor, 1975). Dentro de la hipótesis del lenguaje del pensamiento el contenido y significado derivan del significado de los elementos del lenguaje del pensamiento, la razón de esto es que hay una correlación entre los símbolos y el mundo que dotan de significado a los símbolos mentales. Se podría pensar como relaciones causales, de la misma manera que condiciones de verdad, como, por ejemplo, x causa el símbolo “x” y por ello “x” es verdadero. Esto solamente tomándolo como una relación causal y evidentemente, sin tomar en cuenta ciertos problemas de corte epistemológico como el hecho de que los sentidos podrían engañarnos. Pero explica Fodor que “T es condición de verdad de una oración mental M si y solo si T, en condiciones ideales” (Fodor, 1975).

La idea detrás de la hipótesis del lenguaje del pensamiento es que hay mecanismos cognitivos que están diseñados para poner oraciones en las cajas de creencias cuando y solamente cuando estas sean verdad, y si los mecanismos están trabajando apropiadamente, entonces estas oraciones aparecerán en la caja de creencias cuando estas sean verdaderas.

Block explica que en el funcionalismo hay una semántica del rol funcional, es decir que el significado de los símbolos internos y externos es la manera en cómo funcionan, las representaciones internas tienen una función en nuestros procesos de decisión, debate, resolución de problemas en nuestro pensamiento y en esto consiste su significado. En este sentido el significado de una representación o un estado mental se parecería mucho a la noción de uso de lenguaje de Wittgenstein, en donde el significado es el uso, pero como para el funcionalismo lo que importa es la función, entonces el significado se define bajo esos términos.

De la misma manera que establecer relaciones de identidad por medio de sistemas de etiquetado, la noción del lenguaje del pensamiento se construye fundada en el principio funcionalista. Desde los cimientos de esta hipótesis el funcionalismo ya está presente, si pensamos que los estados intencionales que dentro del sistema son interpretados simbólicamente, y que a su vez esos símbolos hacen referencia a estados de cosas en el mundo, adquieren su significado por su función dentro de un sistema gracias a una relación isomórfica con el mundo que se da gracias a la causalidad, entonces lo que importa dentro de este tipo de explicaciones no es ni siquiera cómo es que los estados físicos dentro de un cerebro se establezcan, sino la función que tienen dentro del sistema.

La posibilidad de pensar la mente desde una perspectiva computacional está ligada al principio funcionalista ya que en la explicación computacional los inputs, outputs y los estados internos del sistema son independientes del material en donde son implementados. Lo que en realidad importa es la función que posibilita comprender a una computadora como tal gracias a su descripción, y en este sentido, como se ha dicho anteriormente, la noción computacional de la mente tiende a reducir las representaciones o los estados mentales a cómputos. En este sentido, el funcionalismo es un principio antibiológico, ya que la descripción del sistema computacional es independiente del material en donde se implemente.

En el apartado anterior se explicó el principio de realizabilidad múltiple, en el cual podemos implementar un sistema de cómputo en cualquier material, lo que importa es la relación de modelaje con el objeto, esto quiere decir que bien podríamos crear una compuerta Y con circuitos eléctricos, con electro magnetos o con gatos y ratones (Block, 1995). La realización del hardware es irrelevante en una descripción computacional, el modelamiento computacional no necesariamente se da en términos biológicos. El problema que se le presenta a aquellos que pretenden realizar una explicación funcionalista es que los humanos, como todos los animales que presentamos conductas inteligentes, tenemos un nivel de descripción biológico.

La ciencia cognitiva pretende que las explicaciones no sean incompatibles, en todo caso lo que se busca es que ambas explicaciones se complementen para encontrar el “software de la mente humana” y así crear máquinas inteligentes a nuestra imagen computacional (Block, 1995).

## Capítulo II: Los límites de la computación frente a la mente

### Inteligencia e intencionalidad

Desde sus inicios, en la filosofía se ha preguntado por una característica que al parecer es fundamental para comprender la naturaleza humana, a saber, la inteligencia. Esta noción ha sido, sobre todo comprendida bajo el concepto de razón, desde los antiguos griegos hasta la modernidad. Sin embargo, hoy en día es la ciencia cognitiva quién ha tratado a la inteligencia como una clase funcional. Esto quiere decir que puede ser investigada empíricamente. Anteriormente hemos visto la propuesta de definición de Turing a partir de su *Test*, pero su respuesta no ha sido suficiente, ya que se queda en la superficie de las definiciones conductistas. Aunque como apunta González (2015), esto es debatible ya que la estrategia de Turing es preguntarse por la capacidad de una máquina para tener una mente e inteligencia, pero él propone que la imitación es una capacidad importante dentro de su planteamiento y lo que él pretende en realidad es reemplazar la pregunta sobre si las máquinas pueden tener estados mentales, sino que plantea que ciertas capacidades pueden aprenderse por medio de la imitación. Cuando nos preguntamos por la inteligencia como una clase funcional lo

hacemos por las funciones involucradas que la constituyen y debe ser investigada empíricamente, en términos de roles causales. De la misma manera que Mendel estudió la función de los genes en el ámbito biológico, lo que se debe de investigar son las funciones, no sus realizaciones materiales. En el caso de la inteligencia, un tópico de investigación es el rol de la inteligencia en la resolución de problemas, toma de decisiones, etc.

Como hemos visto anteriormente, un modelo que se ha pretendido utilizar para describir las funciones que conforman la inteligencia es el computacional. Bajo esta perspectiva, los procesos inteligentes no son muy distintos que los procesos computacionales en la medida que podemos reproducirlos mediante procesos algorítmicos para llegar a determinados fines, de la misma manera en que asumimos que funcionan los procesos inteligentes a partir del reconocimiento de ciertas funciones.

Por otro lado, otro aspecto que parece fundamental para la comprensión de la mente es su característica intencional. En la filosofía la intencionalidad ha sido comprendida como la capacidad de la mente y de los estados mentales de ser acerca de algo, lo que en inglés se expresaría como “*aboutness*” (Searle, 1999), esta capacidad tiene que ver con la posibilidad de representar, significar cosas propiedades o estados de cosas. Cuando se dice que los estados mentales tienen intencionalidad se entendería que son representaciones mentales o que tienen un contenido.

El psicólogo a quién se le atribuye con mayor frecuencia el uso de la palabra intencionalidad, en el sentido al que nos referimos, es Franz Brentano. Él comienza a reutilizar este término a finales del siglo XIX, y ha sido utilizada para referir a los problemas de la representación tanto en la filosofía de la mente como en la filosofía del lenguaje. Existen discusiones contemporáneas sobre la intencionalidad, y éstas son una parte importante de discusiones sobre la naturaleza de la mente.

La noción de intencionalidad no debe ser comprendida como las intenciones, más bien, tenemos que remitirnos a su etimología para entenderla en el sentido al cual se refería Brentano. La palabra “intencionalidad” viene del latín “*tendere*” y la idea detrás de la intencionalidad es que lo mental se dirige o tiende hacia los objetos. De aquí la idea de que la mente es como un arco que apunta sus flechas a diferentes arcos. En el marco de la filosofía medieval se habla de *intentio* como lo que hoy entenderíamos como conceptos, y probablemente sería una noción más cercana a la de intensión. Sin embargo, no debemos de

confundir el concepto de lo intencional y lo intensional. Cuando hablamos de la intención o lo intencionado hacemos referencia a estados mentales específicos que, a diferencia de las creencias, deseos, esperanzas, etc., desempeñan una función que participa en causar estos estados intencionales. Es decir, la intencionalidad es una característica que se presenta en muchos estados mentales como los anteriores, pero no se identifica con ellos, más bien, estos estados presentan intencionalidad. Por otra parte, la noción de intensionalidad está mucho más relacionada con la distinción de Frege entre sentido y referencia (Frege et al., 1985). De acuerdo con esta distinción, el sentido tiene que ver con la intensionalidad como el sentido de las palabras “criatura cordada” y “criatura renada”, que comparten la misma referencia o extensión, aunque la intensión sea distinta. Brentano, así como muchos otros filósofos, sostiene que la intencionalidad es la marca de lo mental, y este concepto ha jugado un papel central en la tradición filosófica analítica, así como en la tradición fenomenológica.

Brentano sostiene que todo fenómeno mental está caracterizado por una noción que resulta bastante controversial tanto para los analíticos como para los fenomenólogos, la inexistencia de los objetos intencionales. Y esto lo podríamos entender como la referencia a un contenido, o dirección hacia un objeto, pero no debemos malentenderlo como algo que significa la cosa. Todo fenómeno mental tiene un objeto dentro de sí, aunque no lo hacen del mismo modo. Habla de una fase de la intencionalidad llamado “presentación”, durante ella algo es presentado y de ello resultarían ciertas actitudes intencionales, por ejemplo, en el amor algo es amado, en el odio algo es odiado, en el deseo algo es deseado, en el juicio algo es afirmado o negado, etc. La inexistencia intencional es una característica exclusiva y definitiva de los fenómenos mentales, ya que solo son estos los que contienen objetos intencionales (Brentano, 2015). El término resulta bastante complejo, ya que como bien dice Searle, es una palabra bastante desafortunada. El término “intexistence” parece poder traducirse como “intexistencia”, pero en inglés parece tener que ver con “in-existence” que en español sería “en-existencia”. Por ello Pierre Jacob se pregunta si Brentano se refería a los objetos a los cuales la mente es dirigida son internos a la mente, o si se refería a la mente que puede ser dirigida a objetos no existentes, o si se refería a ambas (Jacob, 2019). Existen discusiones bastante extensas con respecto al tema de la intencionalidad, sin embargo, para fines de esta tesis solamente mencionaré las tres tesis de Brentano que Jacob rescata, y que serán importantes en los próximos capítulos:



- Primera tesis: Es constitutivo del fenómeno de la intencionalidad, como se exhibe por los estados mentales como amar, odiar, desear, creer, juzgar, percibir, esperar y muchos otros, que estos estados mentales estén dirigidos hacia cosas que son distintas de ellos.
- Segunda tesis: Es característico de los objetos hacia los que la mente es dirigida gracias a la intencionalidad que tengan inexistencia intencional.
- Tercera tesis: Todos y solamente los estados mentales exhiben intencionalidad. (Jacob, 2019)

En las dos primeras tesis lo que tenemos es una relación en la cual los estados mentales son dirigidos a los objetos que tienen inexistencia intencional, por lo que son dos tesis, que, por decirlo de alguna manera, están casadas. Pero la tercera tesis parece muy importante, sobre todo cuando se pretende establecer una distinción entre la cognición humana y el procesamiento computacional, ya que parecería que en ella encontramos a la intencionalidad como el criterio para determinar cuando un estado es mental. Es decir que, al menos para la tradición que sigue a Brentano, para que un estado sea considerado como estado mental debe de ser acerca de algo.

Una de las preguntas más importantes para la ciencia cognitiva, en tanto que pretende dar cuenta de la mente, y sus características como la inteligencia y la intencionalidad, es la de explicar cuando un estado es mental. El problema que se les presenta en primera instancia es que la explicación que se requeriría es una que tome en cuenta las relaciones causales, es decir, que una explicación que apele a otro tipo de sustancias inmateriales no sería suficiente. La perspectiva ontológica que se adopta es la del materialismo. Willard Van Orman Quine sostiene la visión lingüística de la intencionalidad (Quine, 2013), lo que él pretende es llevar la discusión al plano lingüístico, para él las oraciones que proferimos en los lenguajes naturales tienen un significado, y en tanto que lo tienen puede tener la misma forma que los estados mentales y por ello son dirigidas hacia otras cosas que no son ellas, de las cuales algunas no necesitan existir espaciotemporalmente. Aquí lo que probablemente se debería de apuntar es que las preferencias no son cosas mentales, pero sí exhiben un tipo de intencionalidad. Searle establece una distinción entre intencionalidad primitiva e intencionalidad derivada (Searle, 1999), la segunda se podría comprender como un estado

intencional degradado que depende de las oraciones, mientras que las primitivas sí tendrían que ver con estados mentales. Esto se podría explicar de la siguiente manera: Un ser humano común y corriente tiene la capacidad de tener pensamientos sobre las cosas, estos pensamientos tendrían que ser categorizados como estados mentales que tienen una intencionalidad primitiva ya que tienen una relación causal con los objetos a los cuales son dirigidos, podríamos pensar que cuando un objeto espaciotemporal hiere los sentidos, entonces entra a modo de input a un sistema humano. Por otro lado, cuando la misma persona decide plasmar sus pensamientos en una hoja de papel con tinta lo hace codificando sintácticamente una serie de símbolos de manera tal que si otro humano los lee, los puede entender. En este sentido, las palabras en el papel tienen un contenido intencional, pero es una intencionalidad derivada y su existencia depende de las oraciones que funcionan como transportadoras de contenido intencional, aunque no se identifican con él.

Siguiendo a Quine, las oraciones de lenguajes naturales no tienen un significado intrínseco, de la misma manera que las preferencias de las oraciones tampoco lo tienen. Es decir que, las oraciones proferidas en lenguajes naturales no tienen un significado a menos que éste les sea conferido por personas que las usen para expresar sus pensamientos y comunicarlos a otros.

Ahora bien, una pregunta obligada que se debe de hacer dado que nuestras preferencias en un lenguaje natural tienen intencionalidad derivada, es: ¿En dónde reside la intencionalidad primitiva bajo un esquema fisicalista? Como hemos visto en los apartados anteriores, Jerry Fodor postula la noción de lenguaje del pensamiento, éste consiste en símbolos mentales con propiedades semánticas y sintácticas, y entonces, posiblemente, las propiedades semánticas de los símbolos mentales son los portadores de la intencionalidad original o, en términos de Searle, primitiva. Si aceptamos una ontología fisicalista y la tercera tesis de Brentano, entonces tenemos que dar un criterio para lo mental en términos causales. Además, si establecemos una relación entre estados mentales con estados neuronales, debemos preguntarnos si algo no mental exhibe intencionalidad primitiva. Esto porque el tipo de explicación fisicalista que daría cuenta de los estados mentales como estados neuronales tendría que explicar en términos no mentales lo mental.

El problema con el que nos encontramos aquí es, en términos de Fodor, que “la semántica se muestra permanentemente recalcitrante a la integración al orden natural dado que en una

ontología fisicalista, la intencionalidad o las propiedades semánticas no pueden ser características fundamentales del mundo” (J. A. Fodor, 1979). Es decir que el problema al que se enfrenta el fisicalismo es mostrar como un sistema completamente físico puede exhibir estados intencionales.

El problema con el que se encuentran aquellos que pretenden explicar la intencionalidad, es su emergencia. Es decir, se tiene que dar cuenta de cómo es que, a partir de un estado no mental, un estado no intencional, se puede dar tal cosa como la intencionalidad. Para ello se deben hacer las siguientes preguntas: ¿Todo estado mental es intencional? ¿Puede haber algo que no sea mental y que presente intencionalidad?

Como hemos visto antes, sí se podría argumentar que hay cosas no mentales exhiben intencionalidad, como un termómetro o la aguja de la gasolina en un auto. Sin embargo, parece que para dar cuenta de la mente no debemos de apelar a la intencionalidad derivada, sino que tenemos que enfocarnos en la manera en que se genera la intencionalidad primitiva.

Una manera de abordar este problema es la ya mencionada noción del cerebro como un motor sintáctico que conduce un motor semántico. Desde esta perspectiva sí se parte de que un sistema inteligente tendría que ver con la capacidad de éste para presentar ciertos comportamientos que nos permitan reconocer su inteligencia, bajo la teoría computacional de la mente esta conducta podría ser explicada por un sistema determinado sintácticamente que ejecuta programas siguiendo directivas de manera recursiva hasta llegar al resultado deseado. Es decir que, presuntamente la inteligencia estaría relacionada con la capacidad de resolución de problemas, a modo de razón práctica, se establecen pasos para llegar a determinados fines. Lo que haría inteligente a un sistema es lo que puede hacer, lo que tiene capacidad de realizar. Parecería que el modelo de análisis funcional podría explicar procesos inteligentes. Pero, ¿podría explicar un sistema intencional?

Ya analizamos que las palabras escritas en un papel tienen un tipo de intencionalidad derivada y, además, no presentan comportamiento inteligente. Podemos asumir que los contenidos intencionales primitivos son el contenido intencional que las representaciones de un sistema tienen para sí mismo. Una cualidad de la que habla Block con respecto a la intencionalidad es que depende de su naturaleza causal, debe de haber una historia que haga que los estados mentales representen el mundo, esto es, que sean acerca de algo (Block, 1995b). Y, por ende, la intencionalidad tiene un requerimiento orientado hacia el pasado

mientas que la inteligencia tendría una capacidad orientada hacia el futuro. Un ejemplo de esto es el experimento mental del hombre del pantano:

Imagine que un evento con una mínima posibilidad de ocurrencia (pero de manera relevante no cero): Dados los movimientos aleatorios de las partículas de un pantano, éstas se ensamblan y, por azar, crean una réplica que es molécula por molécula igual a su cerebro. El cerebro-pantano es supuestamente inteligente porque tiene muchas de las mismas capacidades que su cerebro tiene. Si lo conectáramos a los inputs y outputs correctos, y le diéramos un problema aritmético, obtendríamos una respuesta inteligente. Pero hay razones para negar que éste tenga los estados intencionales que usted tiene, es más, para negar que tenga estados intencionales en lo absoluto.

La principal razón para argumentar que el hombre del pantano no tiene estados intencionales sería que, en primera instancia, no ha sido conectado a dispositivos de input, esto quiere decir que nunca ha tenido algún tipo de relación con el mundo que le permita tener estados que sean sobre algo. Podríamos, incluso, pensar que, ya que el hombre del pantano tiene estados físicos idénticos a los míos, sus estados neuronales podrían tener el mismo contenido intencional por su forma sintáctica, entonces sería posible asumir que tiene el pensamiento de que Lionel Messi ganó el Balón de Oro. Dado que los estados físicos del cerebro del hombre del pantano como los míos son idénticos a nivel molecular, sería factible pensar que su contenido fenoménico también es igual, lo que sucede en nuestros cerebros, en términos causales, es lo mismo. Ahora bien, Block argumentaría que sí hay una diferencia importante: Yo, a diferencia del hombre del pantano, sí sé quién es Messi y sé qué significa ganar el Balón de Oro, aunque no de primera mano. Por su parte, mi copia no tiene una historia causal con respecto a Messi, ni al Balón de Oro, incluso sería cuestionable que sepa qué es el verbo ganar. El hombre del pantano no tiene un referente que haya causado esas representaciones, de hecho, sería problemático llamarlas representaciones para un sistema que no tiene nada que representar, además de que iría en contra de las primeras dos tesis de Brentano, el hombre del pantano no puede tener intencionalidad porque no está dirigido a algo que tenga existencia intencional, y a pesar de que Messi y el Balón de Oro sí tienen una existencia en el mundo espaciotemporal, no contarían como inexistentes

intencionalmente porque no tienen una relación con el contenido simbólico que tendría el hombre del pantano.

Mediante este ejemplo lo que se pretende mostrar es que puede haber sistemas con capacidades inteligentes, pero eso no implica que sean sistemas intencionales. Además, la inteligencia, desde la perspectiva computacional de la mente, dependería de la estructura sintáctica de determinados estados, que no necesariamente son mentales, y que por medio de ellos es posible resolver problemas matemáticos, tomar decisiones, etc. La manera de explicar la intencionalidad bajo esta perspectiva computacional es con el argumento de que el cerebro es un motor sintáctico que lleva a un motor semántico, en el apartado anterior se habló de que en este argumento se pretende explicar la intencionalidad mediante el isomorfismo que tiene la función sintáctica con la función semántica ya que hay aspectos físicos que pueden ser interpretados simbólicamente. La idea es que tenemos estructuras simbólicas en nuestro cerebro, y que la naturaleza se ha ocupado de que haya correlaciones entre las interacciones causales de estas estructuras y las relaciones racionales de los significados de las estructuras simbólicas (Block, 1995b). Lo que sucede, de acuerdo con este planteamiento, es que los procesadores primitivos “conocen” las formas sintácticas de los símbolos que procesan, sin embargo, son ciegos al sentido, aunque controlan procesos que “tienen sentido”.

La ciencia cognitiva busca dar cuenta de la mente mediante una perspectiva computacional que está fundada en el análisis funcional en la cual la intencionalidad se justificaría mediante el isomorfismo que hay entre una función sintáctica, en la cual los procesos sintácticos están codificados en los procesadores primitivos que a su vez crean procesos complejos, y en una función semántica, en la cual los símbolos como los números tienen la misma forma que los numerales y por ello su procesamiento sintáctico cuenta también como procesamiento semántico. Sin embargo, los procesadores primitivos, ni las funciones sintácticas comprenden el contenido intencional de la función semántica. A lo sumo, la función sintáctica cumple el rol de transportar mediante símbolos un contenido intencional.

La razón por la que sostendría que hay un contenido intencional es porque, siguiendo a Russell en su artículo “On denoting” (Russell, 1905), él acepta que los nombres del lenguaje natural como “Pegaso” no son nombres propios lógicos. Él sostiene la visión epistemológica

en la cual uno no puede usar en el pensamiento un nombre propio lógico que se refiera a algo concreto. Tampoco podría considerar que un pensamiento singular sobre eso es genuino. Si uno está directamente relacionado con un objeto, entonces en vez de eso uno debe de formar un pensamiento general que no es sobre ningún individuo particular. Desde la perspectiva de Russell los nombres de los lenguajes naturales no son nombres genuinos, sino que son descripciones definidas disfrazadas o abreviadas. Su función lógica es la de una descripción definida. Por ejemplo, el nombre de Pegaso solamente sería una abreviatura de la función lógica que tendría “el caballo alado”. Evidentemente este tipo de análisis se aplicaría al reino lingüístico, sin embargo, parecería que si uno puede asumir que los conceptos que son codificados, por ejemplo, en un papel con tinta, tendrían un contenido en el sentido de que hay una construcción sintáctica en la cual hay símbolos plasmados en una hoja de papel, entonces hay objetos no mentales que tienen un contenido intencional derivado y codificado en forma de contenido, asociando la intencionalidad por medio de la propiedad lógica de la connotación, como el sentido que está codificado mediante símbolos con un aparato sintáctico, con la intencionalidad derivada. Ahora bien, algo que se podría argumentar es que, si no hay un hablante del lenguaje natural en el cual está escrito el mensaje, o al menos alguien con las herramientas y capacidades cognitivas que le permitieran decodificarlo, el mensaje fallará en su función de ser interpretado como un portador de sentido de la misma manera que muchos jeroglíficos egipcios lo hicieron hasta que se encontró la Piedra Rosetta.

El problema que se plantea es que sí podríamos contemplar el hecho de que hay objetos con intencionalidad derivada y que hay sistemas inteligentes que no necesariamente tienen intencionalidad primitiva. Pero el tema es: ¿Cuál es el criterio para que algo sea considerado mental?

Searle argumenta que el rol primario de la mente es el de relacionarnos de ciertas maneras con nuestro ambiente, y especialmente con otras personas. Los estados subjetivos nos relacionan con el resto del mundo y el nombre de esa relación es intencionalidad, comprendiéndola como las distintas formas en que la mente puede ser dirigida hacia algo, o ser sobre algo, ya sean objetos o estados de cosas en el mundo (Searle, 1999). Las percepciones, creencias, deseos e intenciones se expresan como actitudes proposicionales y como tales son estados mentales con intencionalidad. Son sobre algo o representan objetos o estados de cosas y tienen un formato psicológico. Searle apunta que hay una relación

importante entre conciencia e intencionalidad. Él argumenta que no todos los estados intencionales son conscientes y no todos los estados conscientes son intencionales.

La relación es esta: Los estados cerebrales que no son conscientes pueden ser comprendidos como estados mentales solo en el grado que podamos comprenderlos como capaces, en principio, de dar lugar a estados conscientes. Mi creencia, por ejemplo, de que Clinton es presidente de los Estados Unidos puede ser consciente o inconsciente. Yo puedo, por ejemplo, ser veraz al decir que tengo esa creencia incluso cuando estoy dormido. ¿Pero qué hecho corresponde a esa afirmación cuando estoy completamente inconsciente? Los únicos hechos que realmente existen en el acto son hechos que implican estados de mi cerebro que son describibles en términos puramente neurobiológicos (Searle, 1999).

Searle argumenta que lo único que podría hacer que un estado sea mental es que sean en principio capaces de causar otro estado de forma consciente, por ejemplo, un estado inconsciente es el tipo de cosa que podría ser consciente. Es por eso que distingue entre los estados inconscientes del cerebro, como la segregación de ciertos neurotransmisores, de los estados mentales no conscientes que son realizados en el cerebro. Él sostiene que solamente los segundos tienen la capacidad de ser causar estados mentales. Hace una alegoría con respecto a una computadora, dice que podemos imaginar al cerebro como una computadora, cuando la apagamos todas las palabras e imágenes en la pantalla desaparecen, pero no dejan de existir. Más bien, son guardadas en el disco duro en forma de una cinta magnética, pero no en forma de imágenes y palabras. El hecho de que sigan siendo palabras e imágenes está constituido en que esas cintas magnéticas puedan ser convertidas en palabras e imágenes cuando encendamos la computadora. Lo mismo sucedería con los estados mentales, que aunque no son palabras e imágenes como las de una computadora, son parecidos a estos estados de la computadora cuando no están en la pantalla. Estos estados mentales tendrían una forma diferente, pero siguen siendo estados mentales no conscientes que son capaces de causar estados mentales conscientes. Sin embargo, los estados mentales no conscientes no pierden su característica mental en virtud de que pueden convertirse en estados conscientes.

Searle sostiene que lo que él llama “principio de conexión”, de acuerdo con el cual a menos que un estado mental esté disponible para la conciencia, no califica como un estado

genuinamente mental. Para él, la visión en donde la disponibilidad para la conciencia es el verdadero criterio para lo mental implica que los estados y procesos que son investigados por la ciencia cognitiva y que no son disponibles para la conciencia fallarán para calificar como estados mentales genuinos. Sin embargo, a pesar de que parecería que Searle no está de acuerdo con la tercera tesis de Brentano, la intencionalidad sería un factor importante de la conciencia, ya que para él una posible explicación de ésta es mediante la naturalización de la conciencia mediante una teoría biológica, tema que se retomará más adelante.

En conclusión, podemos comprender la intencionalidad como la capacidad de representar o de ser sobre algo. Ésta, a la par que la inteligencia es una característica primordial de la mente, y si lo que se pretende es comprenderla, entonces se debe de abordar con la misma seriedad que se ha abordado la noción de inteligencia. Una de las aproximaciones contemporáneas más importantes que trata la intencionalidad es la ciencia cognitiva y lo hace de la misma manera que aborda a la inteligencia, mediante el análisis funcional. Sin embargo, parece ser un camino arduo, ya que la intencionalidad presenta características distintas a la inteligencia en el sentido de que la teoría computacional de la mente parece no poder abarcar por completo todas las nociones de intencionalidad, y en específico parece no dar cuenta de la intencionalidad primitiva mediante el recurso del argumento que el cerebro es un motor sintáctico que lleva un motor semántico, ya que parece que esta aproximación, a lo sumo, solamente podría dar cuenta de la intencionalidad derivada que es interpretada semánticamente mediante procesos sintácticos que son ciegos al sentido de los símbolos que transporta. Además, el problema de determinar cuando un estado intencional, que se supone es la marca de lo mental, resulta bastante problemático, porque al parecer no se puede dar cuenta de la intencionalidad mediante una perspectiva computacional, que sí daría cuenta de un sistema inteligente, pero no intencional. Es por ello que Searle deja de apelar a la intencionalidad como la marca de lo mental y se enfoca en los estados conscientes causados por el cerebro.

## El Argumento de la Habitación China de Searle

El argumento de la habitación china de Searle es uno de los temas más discutidos hoy en día en la filosofía de la mente y en la ciencia cognitiva. Éste aparece por primera vez en el



artículo *Minds, Brains and Programs* en la revista *The Behavioral and Brain Sciences* en 1980 (Searle, 1980). El problema que se plantea es si un programa computacional que parece comprender el lenguaje puede, de hecho, producir comprensión verdadera de historias. El experimento de Searle se inspira en el Test de Turing. Turing se pregunta por un criterio que nos permita definir la inteligencia. Searle, por su parte, no solo se pregunta por una definición de inteligencia, sino que trata de comprenderla no solamente por estructuras formales, sino que también por lo que implicaría el entendimiento lingüístico. Asimismo, él sostendría que las computadoras no gozan de este poder porque su construcción funcional simplemente depende de un aparato sintáctico que les permite manipular símbolos, pero no comprenden los símbolos que procesan porque son ciegos a los contenidos semánticos. A diferencia de una computadora digital, las mentes tendrían que ser el resultado de procesos biológicos, y en el mejor de los casos las computadoras podrían simular algunos de estos procesos. La diferencia entre la inteligencia relacionada con una función sintáctica y la intencionalidad con la función semántica es indispensable para comprender este experimento mental, la razón de esto está en que la noción computacional del motor sintáctico que lleva un motor semántico es fuertemente cuestionada, si no es que desechada, por este argumento.

Este experimento mental aborda un problema filosófico actual y que, con los avances en la tecnología de Inteligencia Artificial, cada vez resultaría más relevante. Hoy en día los seres humanos convivimos con muchos dispositivos que ocupan la denominada Inteligencia Artificial, por ejemplo, uno puede entrar a un portal de ajedrez en el cual se puede jugar contra una computadora, en los videojuegos con modalidad PvC (*Player versus Computer*) los jugadores compiten directamente contra la computadora en juegos sumamente complejos como juegos de guerra, fútbol, estrategia, etc., compañías como Apple o Amazon tienen programas computacionales como Siri o Alexa que no solamente cumplen ciertas funciones de *chatbots* sino que tienen acceso a una serie de dispositivos e interactúan con uno de una manera muy parecida que lo haría un asistente personal, la compañía Tesla y Google llevan ya varios años desarrollando automóviles que se manejan autónomamente, o incluso podemos pensar en los auto correctores en los teclados de nuestros *smartphones* que completan nuestras oraciones gracias a un algoritmo que a partir de nuestros hábitos de escritura pueden predecir cuales palabras utilizaremos. Podríamos suponer que para realizar todos estos procesos se requiere la comprensión y la inteligencia, incluso hay quienes

sostendrían que las computadoras ya comprenden el lenguaje natural. Turing, por ejemplo, sostendría que sí, pero según su criterio conductista, muchas computadoras ya tendrían inteligencia. Muchas personas podrían incluso pensar que las computadoras, de hecho, piensan y, gracias a todos estos comportamientos anteriores, se podría concebir que también comprenden el lenguaje natural. Schank, por ejemplo, desarrolla una técnica llamada “La representación conceptual”, en la cual usaba scripts para representar relaciones conceptuales que, según él, permitían que la IA comprendieran las oraciones que se le presentaban. Él afirma que la IA puede comprender historias sobre temas que conoce. Su idea en términos generales era la de hacer traducciones por medio de un programa computacional, este programa tendría codificado el sentido del texto de manera sintáctica, es decir, con independencia del idioma en el que los textos estuvieran escritos. De este modo se necesitarían dos programas, uno que analice el lenguaje natural que lo interpretaría como una representación sintáctica de su sentido, y por otra parte se necesitaría un generador que recompusiera las estructuras sintácticas de nuevo en el lenguaje natural (Schank & Abelson, 1977). La aproximación de Schank y Abelson era el principal objetivo de la crítica de Searle.

La pregunta por la inteligencia de las computadoras o sistemas más análogos, por decirlo de alguna manera, no es nueva. El mismo Descartes se pregunta por la inteligencia de las máquinas y las cuestiona apelando al carácter universal de la razón.

(...) mientras que las hubiere (máquinas) que semejasen a nuestros cuerpos e imitasen nuestras acciones, cuanto fuere moralmente posible, siempre tendríamos dos medios muy ciertos para reconocer que no por eso son hombres verdaderos; y es el primero, que nunca podrían hacer uso de palabras ni otros signos, componiéndolos, como hacemos nosotros, para declarar nuestros pensamientos a los demás, pues si bien se puede concebir que una máquina esté de tal modo hecha, que profiera palabras y hasta las profiera a propósito de acciones corporales que causen alguna alteración en sus órganos, como, verbi gratia, si se le toca en una parte, que pregunte lo que se le quiere decir, y si en otra, que grite que se le hace daño, y otras cosas por el mismo estilo, sin embargo, no se concibe que ordene en varios modos las palabras para contestar al sentido de todo lo que en su presencia se diga, como pueden hacerlo aun los más estúpidos de entre los hombres; y es el segundo que, aun cuando hicieran cosas tan bien y acaso mejor ninguno de nosotros, no dejarían de fallar en otras, por donde se descubriría que no obran por conocimiento, sino sólo por la

disposición de sus órganos, pues mientras que la razón es un instrumento universal, que puede servir en todas las coyunturas, esos órganos, en cambio, necesitan una particular disposición para cada acción particular (Descartes, 2006).

Parecería que Descartes se mostraría bastante escéptico al comportamiento inteligente de las máquinas porque, y con razón, argumentaría que cualquier disposición física que tengan estaría completamente determinada, al contrario de la razón o la mente que tendría un carácter que él comprendería como indeterminado.

Asimismo, como apunta Cole en su artículo *The Chinese Room Argument* (Cole, 2020), existieron algunos otros precedentes que se preguntan por la inteligencia de las máquinas. Uno de ellos data del siglo XVII con el molino de Leibniz. En este experimento mental Leibniz imagina que hubiera una máquina construida para pensar, sentir y tener percepción. Esta máquina sería del tamaño de un molino, ya que en su experimento él pretende entrar a él, cuando se examina su interior encontraríamos partes que funcionan una sobre otra, pero las partes no explicarían la percepción. Lo que Leibniz hace es contrastar el comportamiento de la máquina, que podría parecer ser el producto de un tipo de pensamiento consciente con la manera en que opera internamente. Él apunta que las operaciones mecánicas internas solamente son partes que se mueven de un lugar a otro, pero que no hay nada que sea consciente o que pueda explicar el pensamiento humano. Para Leibniz los estados físicos no son suficientes ni constitutivos para los estados mentales.

Otro precedente del experimento de la habitación china de Searle es uno más contemporáneo y se lo debemos a Alan Turing. Él imagina una máquina de papel que tendría características muy parecidas a la Máquina de Turing, en tanto que sería completamente algorítmica. La diferencia entre la Máquina de Turing y su Máquina de Papel, es que la segunda estaría implementada por un humano, aunque compartiría las mismas características funcionales que tendría una máquina teórica. La máquina de papel es un programa escrito en papel a modo de instrucciones en un lenguaje natural, el humano que la implementa simplemente debe de seguir las instrucciones sin que se le requiera ningún tipo de ejercicio mental complejo o introspectivo. El experimento consiste en que un hombre está solo en una habitación dónde se le pasan tarjetas por debajo de la puerta y él responde con instrucciones cosas como “N-KB3”, uno ni siquiera tendría que saber que está jugando un juego de ajedrez contra alguien afuera (Cole, 2020). Alguien que estuviera fuera solamente vería como se

mueven las piezas en función de los outputs que genera el hombre siguiendo el programa en la habitación y el jugador que está afuera, entonces podríamos pensar que la persona dentro sabe jugar ajedrez.

Finalmente, un precedente importante es el experimento mental de la Nación China. Este argumento es presentado en el artículo *Troubles with Functionalism* de Ned Block (Block, 1978). En él se debe de suponer que a todos los ciudadanos chinos se les entrega una lista de números telefónicos y se establece un horario en el cual se implementará un programa a gran escala. Los ciudadanos que son designados como inputs inician un proceso de llamado a todos aquellos que están en su lista. Cuando el teléfono de cualquier ciudadano suene entonces ellos deben de llamar a los números de sus respectivas listas, quienes una vez recibida la llamada harán lo mismo. No es necesario que se intercambie ningún tipo de información en la llamada, lo único que se requiere es generar un patrón de llamadas. Las listas de números a llamar estarían construidas de manera tal que implementen los mismos patrones de activación que ocurren entre las neuronas del cerebro de un ser humano cuando está en un estado mental como el dolor. Lo que argumentaría Block es que las llamadas podrían tener el mismo rol funcional que las neuronas provocan entre ellas.

El interés de Block en este experimento mental estaba enfocado en los *qualia*, es decir, la característica cualitativa que tendrían algunos estados mentales conscientes, lo que él se preguntaba es que, si es plausible sostener que la población china pudiese tener dolor colectivamente mientras que ninguno de los ciudadanos, de hecho, lo experimentara. Sin embargo, este experimento mental puede funcionar para preguntarse sobre cualquier estado o proceso mental, incluyendo aquellos que lo hacen por la comprensión lingüística.

Ahora bien, el argumento de la habitación china es un argumento que no está en contra de toda la Inteligencia Artificial, de hecho, está enfocado solamente a una noción, la de la IA fuerte. Hay una distinción entre la inteligencia artificial débil y fuerte. La primera no pretende sostener que la IA pueda comprender el lenguaje, se entiende como el uso de programas como herramientas efectivas que sirven para probar hipótesis científicas sobre la cognición humana. La segunda refiere al tipo de inteligencia artificial que sostiene Schanck y Abelson (Schanck & Abelson, 1977), ésta sostendría que una computadora, SAM, ha sido programada apropiadamente puede tener comprensión del lenguaje. Searle analiza la noción de

comprensión y propone que esta facultad tiene que ver con la capacidad de responder preguntas sobre historias, incluso si en ellas no tiene la información completa.

El argumento de la habitación china tiene el mismo sentido que sus predecesores. A grandes rasgos, Searle se imagina siguiendo un programa de procesamiento simbólico escrito en inglés, y en este sentido lo que estaría implementando también sería lo que Turing llamaría una máquina de papel. El hablante de inglés estaría encerrado en una habitación siguiendo instrucciones en inglés, un lenguaje natural, que le dirían como manipular símbolos en chino enteramente por su forma. El experimento es muy parecido al Test de Turing en el sentido de que un hablante de chino que esté afuera de la habitación podría pensar que la persona que está dentro, de hecho, hablaría chino. Sin embargo, el humano dentro solamente aparenta comprender chino al seguir las instrucciones de manipulación de símbolos, pero no se acerca ni remotamente a comprender ese lenguaje. Dado que una computadora solamente hace lo que el humano, es decir, manipular símbolos basándose solo en una función sintáctica, ninguna computadora, solamente por ejecutar un programa, genuinamente comprendería el chino (Searle, 1980).

González (2016) argumenta que la intención del argumento de la habitación china es la de socavar la IA fuerte mostrando que las computadoras bien programadas no pueden comprender historias, correr un programa no provee una condición suficiente para explicar la comprensión del lenguaje y que los estados mentales no emergen de programas que son ejecutados por computadoras.

Dentro de la habitación china se asume que el manual de manipulación simbólica está bien escrito y que el humano dentro puede correlacionar los símbolos de manera rápida y efectiva para poder enviar los outputs correctos, él siempre podría contestar las preguntas sin saber absolutamente nada de lo que los símbolos chinos dicen. El hablante de chino de fuera podría pensar que hay alguien que entiende chino a la perfección dentro de la habitación. La conclusión del argumento, en este sentido, sería que manipular símbolos con un libro de reglas, de la misma manera en que un programa procesa símbolos al ingresarlos en forma de inputs y arrojando sus conclusiones en forma de outputs, no se puede dar cuenta de la comprensión de historias. Esto quiere decir que, de hecho, los programas computacionales no explican la comprensión ni causan estados mentales, dado que ejecutar un programa no provee una condición suficiente para elucidar la comprensión y tales estados mentales.

El argumento de la habitación china ha sido uno de los temas más controversiales dentro de la filosofía de la mente y de la ciencia cognitiva, por ello es un tema ampliamente discutido y se han dado muchas respuestas. González hace una compilación de ellas y explica cuales son los contraargumentos para cada una (González, 2007).

La primera respuesta sería la de los sistemas, en ella se asume que el hombre dentro de la habitación en conjunto con el manual de manipulación simbólica, la base de datos y la correlación pueden comprender chino. Sin embargo, la respuesta que daría Searle es que si la persona puede memorizar los elementos que están dentro de la habitación y realiza la manipulación de símbolos no surgiría la comprensión. Apunta que hay dos subsistemas que no pueden ser igualados: Uno manipula los símbolos chinos a pesar de que es ciego a su contenido semántico, mientras que el otro comprende los símbolos que ya han sido manipulados, este segundo es un subsistema que funciona en inglés y sí comprende los significados. Además, si asumimos esta respuesta, entonces la noción de sistema cognitivo se trivializaría ya que cualquier otro sistema lo sería. Asimismo, argumenta que en el nivel bajo de la descripción el estómago y el hígado implementan programas, pero esto no les da derecho a decir que comprenden. Si así fuera, ¿Cuál sería el carácter distintivo de la mente? Lo que se haría sería negar la parte cognitiva, es decir a los mismos estados mentales.

La respuesta del robot, en esta se supone que hay un programa computacional que opera un robot permitiéndole poder percibir, caminar, moverse, etc. Apelando a las relaciones causales de los símbolos. Sin embargo, Searle contestaría que aun así, el robot no tendría comprensión de lenguaje ya que, si entramos a su interior, no podríamos saber qué significan los símbolos que está procesando.

En el simulador cerebral se supone que hay una inteligencia artificial tal que es capaz de desarrollar un programa que logre simular los patrones neuronales de un cerebro que comprende chino. El programa, o programas corriendo en paralelo, simularían la estructura formal del cerebro chino. Sin embargo, Searle señalaría que este argumento no sería del todo propio de un defensor de la IA fuerte ya que, si están en lo correcto y la inteligencia se puede realizar en materiales no biológicos, entonces no importaría saber cómo funciona el cerebro para poder reproducir una mente. En este sentido, la crítica de Searle también estaría dirigida al funcionalismo apuntando que este principio trivializaría la investigación de las funciones cerebrales y, entonces, sería mejor simplemente apelar a funciones computacionales en vez

de investigar la mente y su relación con el cerebro. Una variante de esta respuesta se asemeja bastante al experimento de la nación china, se imagina a un hombre que opera tuberías de agua y válvulas dentro de un cerebro que comprende chino, sin embargo, esto no haría una diferencia ya que ese proceso solamente imitaría los inputs y outputs en chino que ocurren en el cerebro, pero no daría cuenta de ninguna comprensión lingüística.

La respuesta de la réplica combinada consiste en conectar las tres respuestas anteriores, ya que juntas serían más convincentes. Parecería que se le podría atribuir intencionalidad al robot gracias a la apariencia y a la conducta del robot, que son causadas por un programa. Sin embargo, esto es insuficiente para atribuir intencionalidad. La razón de ello sería que si hay un hombre dentro de la cabeza del robot, no podría intentar mover partes del robot. El robot sería como un maniquí ingenuo en virtud de su programa, y como consecuencia de ello no le podríamos atribuir una vida mental al robot.

La respuesta de las otras mentes sostiene que nosotros sabemos cuando la gente comprende chino o cualquier otro lenguaje por su conducta. Sin embargo, aquí nos encontraríamos con el mismo problema que en el Test de Turing, no podemos establecer un criterio conductista si es que realmente queremos comprender la mente, incluso funcionalmente, no da ningún tipo de margen para una investigación empírica con respecto a los procesos internos que causarían el comportamiento que identificamos como inteligente. Lo que está en juego no es la manera en que las otras personas tienen estados mentales, sino que la pregunta en realidad es: ¿Qué es lo que se necesita para que un estado sea mental?

Se hace la pregunta sobre la posibilidad de que la tecnología pudiera reproducir causalmente una mente con cerebros artificiales en el futuro. El problema de este planteamiento es que la IA fuerte sería completamente trivial, ya que los procesos computacionales como procesos mentales son formalizados mediante el funcionalismo. En principio, nosotros podríamos darle a la máquina la habilidad para entender lenguajes naturales porque nosotros mismos somos máquinas. Pero, argumenta Searle, implementar un programa no es suficiente para que esa hazaña, fundada en procesos sintácticos que es la base de todo programa, ya que no tiene poderes causales. Para que podamos crear una mente tenemos que poder reproducir esos poderes causales que sí tiene el cerebro. Es decir que Searle, en principio, no descartaría la posibilidad futura de crear un cerebro artificial, en principio parecería que no habría ningún tipo de impedimento lógico.

Finalmente, la respuesta de Churchland es que podría haber una red conexionista que esté inspirada en la neurología, esta red podría tener comprensión. Pero Searle contesta que cualquier red conexionista podría correr en una máquina de procesamiento lineal y entonces todas las consecuencias de la habitación china también serían válidas para una máquina que computa en paralelo. La respuesta de Searle ante las redes en paralelo es la del argumento del gimnasio chino, en este se muestra como la instanciación de una red conexionista no es suficiente para producir la comprensión del chino. Este argumento es una extensión del argumento básico de la Habitación China en el cual en vez de haber un solo hombre hay varios angloparlantes. Los hombres están propagados en la habitación como nodos en una red, y están siguiendo el manual en inglés que les dice cuales símbolos pasarse entre ellos. A través de este procedimiento ellos realizarían las mismas computaciones que una red conexionistas que producirían oraciones en chino, sin embargo, ninguno de los hombres realmente comprendería ese idioma (Searle, 1990).

### La intencionalidad como condición necesaria para la comprensión lingüística

El argumento de la habitación trata de socavar la noción de inteligencia artificial fuerte como aquella que puede comprender historias, sin embargo, parece que ningún sistema basado en procesamiento simbólico podría comprender historias sin tener comprensión del lenguaje. Como hemos visto antes, la distinción que se plantea desde la teoría computacional de la mente, entre un sistema inteligente y un sistema intencional, solamente se podría dar cuenta de los comportamientos inteligentes de determinados sistemas que están programados para responder de esas maneras, pero cuando hablamos de dar cuenta de un sistema intencional parece que el tema se vuelve más complejo.

Como hemos visto anteriormente, podemos comprender la intencionalidad como ser sobre algo o *aboutness*. Es la característica que determinaría lo que en la teoría computacional de la mente conocemos como la función semántica, gracias a ella tenemos la capacidad de hacer representaciones de algo.

El argumento de la habitación china de Searle dice que ésta no produce ningún tipo de comprensión, porque, siguiendo la estrategia de tomar en cuenta la distinción entre el motor sintáctico y el motor semántico, el programa que se implementa en la computadora solamente



responde al orden del primer motor, las relaciones de los símbolos que son procesadas obedecen meramente a estructuras formales y sintácticas. Es decir que, las operaciones computacionales son formales y responden a concatenaciones de símbolos, pero no a su significado. Sin embargo, los humanos no operamos de esta manera, nosotros respondemos a los significados de las palabras, no a su estructura sintáctica o su apariencia física. Las mentes tienen estados con significado, es decir, contenidos mentales (Cole, 2020).

Las computadoras manipulan sintácticamente para responder a inputs que se dan en un lenguaje natural, pero no entienden las oraciones que reciben y tampoco a las que entregan en forma de outputs. Searle caracteriza a los programas como algo meramente formal que tiene que ver con una naturaleza sintáctica, mientras que las mentes humanas tienen un contenido mental o semántico. La sintaxis por sí sola no puede ser suficiente o constitutiva para el contenido semántico. Por lo tanto, los programas no son suficientes ni constitutivos para la mente (Searle, 1984). Lo que hace es comprender que dentro de la teoría computacional de la mente hay un espacio abismal entre la función semántica y la función sintáctica, ya que como habíamos visto antes, a pesar de que un proceso computacional pueda realizar manipulación de símbolos, es ciego ante su significado.

Tomemos, por ejemplo, la manera en que nosotros simbolizamos ciertos procesos lógicos. Sobre todo, después de la lógica proposicional la manera en que nosotros formalizamos las proposiciones es sintetizándolas en forma de letras como “p” o “q”. Para realizar una fórmula bien formada, es decir, un proceso lógico que sea lógicamente válido, utilizamos conectivas lógicas que están definidas por su función dentro de la fórmula y son las que norman la estructura formal. Por ejemplo:

En donde:

P simboliza “mi perro tiene pulgas”

Q simboliza “el gato juega con el perro”

R simboliza “el gato tiene pulgas”

S simboliza “tengo que bañar al perro”

T simboliza “tengo que bañar al gato”

$$[(p \cdot q) \supset r] \supset (s \cdot t)$$

Esta fórmula diría “Si mi perro tiene pulgas y el gato juega con el perro, entonces el gato tiene pulgas. Entonces tengo que bañar al perro y tengo que bañar al gato”.

La fórmula tiene una estructura sintáctica válida, pero eso no significa que la fórmula por sí misma pueda comprender las proposiciones que están siendo representadas por “p”, “q”, “r”, “s” y “t”, de hecho, su contenido semántico es irrelevante para los fines formales que implica una fórmula lógica. Nosotros somos los que le asignamos un valor a cada letra para que represente una proposición, la sintaxis que dan las conectivas lógicas solamente permite que los valores de verdad puedan seguir un flujo como se podría comprobar en una tabla de verdad. Lo importante aquí es que somos los que leemos los símbolos y comprendemos las proposiciones que se les atribuyen los que realmente comprenden toda la formulación gracias a su construcción sintáctica como semántica.

Los símbolos formales por sí mismo nunca pueden ser suficientes para los contenidos mentales, porque los símbolos, por definición, no tienen significados (o interpretación) excepto en la medida de que alguien fuera del sistema se los dé (Searle, 1984).

Detractores de Searle como Chalmers (1996) sostendría, al igual que Searle, que un programa por sí solo no puede causar nada, pero él argumenta que la comprensión se podría dar en el ámbito de la implementación, ya que cuando un programa es ejecutado por una computadora entonces se encienden los poderes causales del sistema físico que es la computadora, y una vez que la computadora está ejecutando el programa comenzaría a tener relaciones causales con sus inputs. Además, se podría argumentar, en contra de Searle y siguiendo la lógica de Chalmers, que una computadora, así como un cerebro, son sistemas físicos que no pueden ser reducidos a un programa formal, sino que cuando están corriendo entonces es cuando comenzaría la relación causal con su ambiente. Dennett, otro crítico de Searle, cree que la programación sí es capaz de darle una mente a algo, sin embargo, él si considera el factor material, ya que dice que “es improbable empíricamente que los tipos de programas correctos puedan ejecutarse en otra cosa que no sea un cerebro humano orgánico” (Dennett, 1987).

Además de estas objeciones, podríamos recurrir al experimento de la Nación China, en él los ciudadanos chinos no están compartiendo ningún tipo de información en las llamadas que

hacen, es más, ni siquiera sería relevante que supieran que están implementado la emulación de un estado mental como lo sería el dolor. De la misma manera, los estados físicos de la computadora no sabrían que están manipulando ceros y unos, ni siquiera saben que están utilizando una notación binaria y somos nosotros los humanos los que, de hecho, estamos asignando un valor de 1 a cargas de 7 voltios con el numeral “1” y un valor de 0 a cargas de 4 voltios con el numeral “0”.

Searle distingue entre dos tipos de fenómenos, aquellos que su existencia es dependiente de un observador y los que no necesitan de un observador para existir. Aquellos que son independientes serían fenómenos como los átomos de hidrógeno, las placas tectónicas, virus, árboles y galaxias. Su realidad es independiente de nosotros los observadores. Este tipo de fenómenos son caracterizados como hechos brutos. El otro tipo de fenómenos son dependientes del observador, por ejemplo, el valor del dinero o las nacionalidades, son hechos que dependen del lenguaje. Este tipo de fenómenos son ontológicamente subjetivos y su existencia depende del observador (Searle, 1999).

Searle acepta que la interpretación que se le da a la mente mediante una aproximación computacional, así como la sintaxis, son hechos relativos al observador. “La computación existe solo en relación con algún agente observador quien impone una interpretación computacional a algún fenómeno” (Searle, 2002). Una consecuencia de aceptar esta noción en términos generales es que el concepto de computación terminaría trivializándose. Como vimos en el primer capítulo tanto Block como Copeland estarían en contra de esto y por lo mismo tratan de establecer criterios sobre lo que se debe de comprender, en especial Copeland cuando establece una relación de modelaje entre una computadora y su diseño como tal.

Otra defensa que se hace con respecto al argumento de la habitación china es la de Haugeland, el sostendría que la IA sí podría tener intencionalidad y comprensión de lenguaje, ya que un procesador comprende intrínsecamente los comandos en los programas que ejecuta, y lo hace de acuerdo con las especificaciones. “La única forma en que podemos dar sentido a una computadora como ejecutando un programa es entendiendo que su proceso responde a las prescripciones del programa como significados” (Haugeland, 2003). Esto querría decir que los símbolos y su estructura, que nosotros interpretaríamos como sintaxis, sí tendrían un significado en el sistema.

La habitación china de Searle sí hace una crítica a la teoría computacional de la mente, sobre todo por la distinción entre función sintáctica y función semántica, sin embargo, él apunta que lo que serviría como un criterio para la comprensión del lenguaje tiene que ver con la facultad de la intencionalidad, parecería que ese criterio sí sería suficiente.

A pesar de las diferencias que Searle tiene con Fodor, sobre todo con respecto a la naturaleza de los símbolos como sintaxis en la noción del lenguaje del pensamiento, los dos aceptan que hay una distinción entre la intencionalidad primitiva u original y la intencionalidad derivada. Como vimos anteriormente, la primera se refiere a la relación causal que tendría la mente para que sus estados sean acerca de algo, mientras que la segunda se trata sobre una intencionalidad que tendrían los objetos no mentales, pero que es derivada de los estados que sí lo son, por ejemplo, las palabras escritas en un papel o las preferencias. Esta distinción aplicaría también a las computadoras, ya que los estados computacionales tienen contenido, pero no tienen intencionalidad primitiva.

Si lo que nos interesa es discernir sobre la diferencia entre la cognición humana y los procesos computacionales, entonces parece que la distinción que nos otorga Searle se vuelve crucial. A pesar de que la aproximación computacional de la mente sí tomaría en cuenta la intencionalidad como una función semántica llevada a cabo por una función sintáctica, su teoría de la intencionalidad que por medio del análisis funcional pretende dotar a los símbolos procesados de contenidos semánticos, parece insuficiente porque no explicaría la intencionalidad primitiva, los contenidos simbólicos solamente caerían bajo la categoría de intencionalidad derivada.

Aparentemente el tema de la intencionalidad original sería lo que, siguiendo la línea de Brentano, podría funcionar como un criterio para establecer lo que cuenta como un estado mental. Originalmente, Searle dirige su argumento contra la intencionalidad de una máquina formal y sintáctica, sin embargo, en sus escritos posteriores como en su texto "Intentionality" (Searle, 1983) queda claro que su problema en realidad es la conciencia, ya que, para él, ésta es la condición necesaria para que tengamos intencionalidad primitiva. Esto gracias a que la intencionalidad implica una capacidad representacional, es decir que está dirigida a algo. Searle, en este sentido trata la intencionalidad como una condición que debe ser satisfecha para que haya referencia. Estas condiciones determinarían el tipo de relación histórica que una mente debe de tener para poder ser considerada un sistema intencional y no ser como el

hombre del pantano, o una computadora que no tiene relaciones causales con los contenidos semánticos que procesa. Estas condiciones son lo que permiten que la mente pueda ser dirigida al mundo y son condiciones necesarias para que haya una comprensión sobre la representación que sí tendría la intencionalidad primitiva.

### Capítulo III: La explicación biológica de la mente y del entendimiento humano

#### La conciencia, una característica primaria de la mente

El parteaguas del argumento de la Habitación China de Searle con respecto a la aproximación computacional de la mente radica en la comprensión del lenguaje. A su vez, ésta estaría fundada en la intencionalidad original, o primitiva. Sin embargo, como vimos en la sección pasada, la respuesta de Searle a la presunción de que, a partir de procesos sintácticos que en virtud de su isomorfismo con una función semántica, se podría pensar que un programa computacional puede comprender historias, no solamente está enfocada a la intencionalidad, parecería que su objetivo va más allá de una mera crítica, sino que busca la caracterización de lo que entenderíamos por los estados mentales.

Desde Brentano, Husserl y la tradición fenomenológica, se ha aceptado que la conciencia tiene una naturaleza irreductible, esto quiere decir que no puede ser comprendida mediante su reducción a estados más simples, de la misma manera que el análisis funcional, como en el ejemplo del homúnculo compuesto por homúnculos cada vez más simples, comprende las funciones de los programas complejos por medio de procesadores primitivos.

Para Searle, la característica más esencial de la mente es la conciencia. Él comprende la conciencia como un conjunto de estados de sintiencia o conciencia (como “*awareness*”) que generalmente comienzan cuando despertamos en la mañana y continuamos durante el día hasta que volvemos a dormir (Searle, 1999). Dentro de nuestra vida consciente tenemos experiencias que pueden ser comprendidas como estados conscientes, pero no todos son iguales, el despertar a las siete de la mañana para ir a trabajar no es la misma experiencia, ni se siente igual, que el escuchar una melodía de Bach, realizar una operación matemática, correr un auto en la carretera o fumar un cigarro. A pesar de que estas acciones implican

experiencias conscientes, parecería que no hay un común denominador que las pueda categorizar como una sola cosa. Sin embargo, Searle apunta a que todas estas experiencias cuentan como estados conscientes porque tienen tres características esenciales de la conciencia: Son internas, cualitativas y subjetivas.

En primera instancia son internas en dos sentidos, el primero es bastante literal, un estado consciente, en tanto que lo podemos comprender también como un estado neuronal, está dentro de nosotros, y más específicamente, está dentro de nuestro cerebro. El segundo sentido tiene que ver con la manera en que este tipo de estados están conectados entre sí, cuando pensamos a un estado consciente lo debemos de hacer en una interrelación con otros estados conscientes, solamente pueden existir como elementos de una secuencia de tales estados. Por ejemplo, cuando se me viene a la mente el viaje que hice a Argentina, entonces no solamente estoy pensando en la oración “Yo viajé a Argentina”, esta no significaría nada si no viniera acompañada de todos los demás estados conscientes que implican que alguien viaje a Argentina como ir a comer choripán o visitar el obelisco, por ejemplo. Los estados mentales se relacionan internamente entre ellos en el sentido de que evocar uno de ellos necesariamente vendrá en relación con otros como un sistema de estados que, por medio de la intencionalidad, están en relación con nuestro ambiente.

La segunda característica es que los estados mentales son cualitativos. Este es un punto que ha sido discutido en la historia de la filosofía, por ejemplo, Locke (1999) ilustraría esta característica por medio de la “rojez” que uno experimenta cuando ve un tomate o el sabor dulce que experimentamos cuando probamos una piña. Sin embargo, el carácter cualitativo parecería no estar solamente relacionado con la sensibilidad como lo sería la “rojez” del rojo o el dulzor de un caramelo, sino que es algo que está presente en todas nuestras experiencias, por ejemplo, el pensar un problema filosófico, o al desear algo. Para filósofos como Thomas Nagel en su famoso artículo “*What is like to be a bat?*”, esta característica es la que marcaría un criterio para determinar cuándo algunas criaturas son realmente conscientes. Si un organismo percibe y responde de manera adecuada a su ambiente, pero carece de los *qualia*, entonces sería muy difícil asumir que es un ser consciente, tal vez en un sentido sumamente laxo sería posible o, en todo caso, de una manera casi metafórica. Él sostiene que para cada estado consciente hay algo que es como estar en ese estado consciente (Nagel, 1974). Los estados conscientes son cualitativos en el sentido de que para cada estado consciente hay una

determinada manera en que se siente, tiene un cierto carácter cualitativo, por ejemplo, hay algo que es como tomar vino, pero es distinto a escuchar música. Apunta Searle que, no hay nada como ser algo así como un árbol o como una casa porque ese tipo de entidades no tienen conciencia (Searle, 1999). Una manera de abordar este carácter de los estados conscientes es la postulación de los *qualia*. Todos sabemos que los seres conscientes tenemos una amplia gama de sentimientos y experiencias como cuando sentimos texturas, sentimos sonidos, olores o dolores, incluso cuando vemos los colores, o cuando nos sentimos felices o iracundos. En todos estos casos somos sujetos de estados mentales con un carácter cualitativo bastante distintivo. Hay algo que se siente de una manera para nosotros en cada estado, una suerte de distintivo fenoménico. Generalmente se utiliza el término “*qualia*”, que tiene que ver con una cualidad singular, para hacer referencia a los aspectos fenoménicos de nuestras vidas conscientes (Peirce, 2020). El carácter cualitativo de los estados conscientes es tan importante que ha sido una de las discusiones más amplias en la filosofía, justamente porque es una noción central sobre la comprensión de la conciencia.

Finalmente, la característica más importante de los estados conscientes es que son subjetivos. Para Nagel los hechos sobre cómo es ser un murciélago son subjetivos en un sentido importante, porque solamente tienen sentido desde el punto de vista de este animal. Solamente las criaturas que son capaces de tener ese tipo de experiencias podrían comprender como es ser ellas (Nagel, 1974). Esto tendría que ver con un problema epistemológico sobre la autoridad de la primera persona sobre sus estados. Aquí se atiende al hecho de que solamente una persona que tiene experiencia, que tendría que ser cualitativa, pero en este sentido habría que prestar especial atención a que es interna como cuando se caracteriza a la conciencia como percepción interna de estados internos (Locke, 1999), tiene primacía para describir sus emociones, sentimientos, deseos, pensamientos, creencias, etc.

Esto quiere decir, en el lenguaje de Searle, que tienen un modo de existencia dependiente del observador, es decir que, siempre son experimentados por un humano o un sujeto animal. Por ello, es que Searle argumenta que tienen una ontología de primera persona (Searle, 1999). Esto quiere decir que solamente existen desde el punto de vista de algún agente u organismo que las experimenta. Esto en contraposición a los modos de existencia de tercera persona, tales como las montañas, los ríos o, para hacerlo más claro, las galaxias que aún no descubrimos.

El hecho de que los estados conscientes sean subjetivos implica que mis estados mentales solamente son accesibles para mí y para nadie más. No podemos negar que cuando caminamos por la casa descalzos y pisamos un Lego el dolor es mucho, y nos sentimos completamente solos en esa experiencia por más que los demás traten de empatizar con nosotros, por supuesto que ellos pueden imaginar el dolor si es que han sufrido eso, pero en ese momento la experiencia se muestra absolutamente subjetiva. Ellos no podrían tener acceso al dolor de la misma manera que yo lo tengo. Searle apunta que ese acceso no es solamente epistémico, sino que el contenido fenoménico de ese estado solamente lo puedo conocer yo mejor que otros y solamente existe para mí. Ahora bien, uno de los problemas que presenta el tema de la autoridad de la primera persona tendría que ver con que, de hecho, a veces nuestros estados mentales bien podrían ser mejor reconocidos por otras personas antes que yo, el mejor ejemplo es el de los celos, yo podría no estar consciente de que de hecho los estoy experimentando, pero mi pareja o mis amigos podrían reconocer ese sentimiento desde su perspectiva y aclarármelo. Por ello el acceso que tenemos a nuestros estados conscientes no es necesariamente epistémico, no se trata de cómo sé sobre ellos, sino que nuestros estados conscientes existen gracias a que son experimentados por nosotros. Searle afirma que la subjetividad de la conciencia la hace irreductible, esto quiere decir que, en tanto que las experiencias que tenemos como estados conscientes no se asemejan al tipo de existencia que tendrían entidades con una ontología de tercera persona como los planetas o las montañas. Y por ello son estados muy difíciles de explicar para la ciencia, y en específico para la ciencia cognitiva que trata de comprenderlos por medio del análisis funcional. La reducción, como hemos visto antes es el método por el cual se explican fenómenos complejos por medio de fenómenos más simples, pero hay distintos tipos de reducción. Searle se enfoca principalmente a dos tipos: la eliminativa y la no eliminativa.

La reducción eliminativa se deshace de los fenómenos al mostrar que realmente no existe, que solamente son una ilusión. Por ejemplo, cuando explicamos la aparición del amanecer y del atardecer, hay un sentido en el cual eliminamos los amaneceres y atardeceres porque mostramos que solamente son ilusiones. El sol realmente no se pone sobre las montañas, la rotación de la tierra en su eje hace que parezca que el sol se pone. (Searle, 1999)



Ahora bien, la reducción no eliminativa funciona de una manera distinta. En ella los objetos a explicar no son mostrados como ilusiones, más bien son explicados por sus componentes. El ejemplo que da Searle es el de la solidez:

La solidez puede ser enteramente explicada causalmente en términos de los movimientos vibratorios de las moléculas en estructuras de celosía. Una vez que las moléculas se están moviendo de esta manera, entonces los objetos son impenetrables por otros objetos. Ellos sostienen otros objetos, y así sucesivamente. La solidez es explicable causalmente en los términos del comportamiento de microelementos, y por esa razón redefinimos la solidez en términos de sus bases causales. (Searle, 1999)

El análisis funcional pretende comprender un fenómeno por medio de la descomposición recursiva de los elementos del objeto que pretende explicar, sin embargo, la tarea parece inconclusa gracias a que aún no hay un consenso con respecto a la naturaleza de la mente. Existen teorías eliminativistas que niegan la existencia de la conciencia o al menos la existencia de sus características más aceptadas. Los más radicales rechazan la existencia de la conciencia, ya que la categorizan como una noción confusa, y afirman que la distinción entre estados conscientes y estados inconscientes no explican la realidad mental. Ellos consideran que la idea de la conciencia está lo suficientemente desviada para merecer la eliminación y el reemplazo por otros conceptos y distinciones que reflejen más la verdadera naturaleza de la mente (Churchland, 1983). Daniel Dennett es conocido por su eliminativismo, él niega la existencia de ciertas propiedades que se le adjudican a los estados mentales conscientes, o al menos que parecen tener, es decir, aquellos estados que son mentales, pero que no implican una representación. Sobre todo, su pugna va en contra de los *qualia* ya que no darían cuenta de cómo es que se generaría una representación, es decir, que serían algo así como un residuo. Lo que Dennett pretende mostrar es que todos los hechos sobre la conciencia pueden ser enmarcados dentro del fisicalismo, sin la necesidad de apelar a los *qualia* o cualquier otro tipo de propiedad mental que no tenga la capacidad de generar representaciones (Dennett, 1992; 1993). Dice que solo una teoría que explique los eventos conscientes en términos de eventos inconscientes podría explicar la conciencia: “To explain is to explain away” (Dennett, 1993).

Searle sostiene que no podemos hacer este tipo de movimientos con la conciencia, no es posible realizar una reducción eliminativista de la conciencia porque el patrón de las reducciones eliminativistas es el de mostrar que los fenómenos reducidos solamente son ilusiones. Pero con respecto a lo que le concierne a la conciencia, la existencia de una “ilusión” es la realidad misma. Es decir, que si me parece que soy consciente, entonces lo estoy. Lo que hace Searle con este contra argumento es un movimiento cartesiano con la misma forma que el “cogito ergo sum”. Si puedo tener una ilusión, entonces esa ilusión no se le puede presentar a nadie más que alguien consciente. La conciencia no se parece a los amaneceres, porque podemos tener la ilusión del sol poniéndose detrás de las montañas cuando en realidad no lo hace. Pero no podemos tener la ilusión de la conciencia si no somos conscientes.

La importancia de la conciencia no solo tiene que ver con las discusiones filosóficas sobre la naturaleza de los estados mentales, sino que también tiene una función en la manera en que nos desempeñamos en nuestro ambiente. El rol evolutivo más primordial de la mente es el de relacionarnos con el mundo, y específicamente con otras personas. Muchas de las acciones que realizamos para sobrevivir como especie requieren de la conciencia, el modo en que los humanos y animales más sofisticados generalmente sobreviven es por medio de la conciencia. Searle argumenta que no podemos decir que la conciencia no cumple con un rol evolutivo, porque es obvio que cumple un gran número de roles. Él diría que la conciencia juega un rol determinante en nuestras vidas, porque sin estados conscientes, por ejemplo, si caemos en un coma y quedamos postrados e indefensos, entonces nos extinguiríamos pronto (Searle, 1999, p.63).

Los estados conscientes tienen causalidad, por ejemplo, el tener una intención o un deseo funcionaría representando el tipo de evento que causa. Por ejemplo, cuando tenemos hambre simplemente vamos a comer algo, el efecto de comer es una representación de su causa, el tener hambre. Esto es a lo que Searle llama “causación intencional”.

Searle tanto en sus libros “Intentionality” y “Mind, Language and Society” (Searle, 1983, 1999) habla de la relación intrínseca que tienen la conciencia y la intencionalidad. Y en este vínculo encontraríamos la clave para determinar un criterio para comprender cual sería la esencia de los estados mentales.

La característica más importante de la conciencia para el propósito de esta discusión es que hay una conexión esencial entre la conciencia y la capacidad que los humanos tenemos para representar objetos y estados de cosas en el mundo para nosotros mismos. Es una característica general de la mayoría, aunque no de todos, los fenómenos conscientes que ellos representan objetos, eventos y estados de cosas en el mundo. De hecho, la característica más importante de la conciencia para los propósitos de esta discusión es que hay una conexión esencial entre la conciencia y la capacidad que, nosotros, los seres humanos tenemos para representar objetos y estados de cosas en el mundo para nosotros mismos. Esta es una característica poseída por las creencias y deseos, esperanzas y miedos, amor y odio, orgullo y vergüenza, así como la percepción y las intenciones. Es una característica que tiene un nombre técnico en la filosofía: Intencionalidad (Searle, 1999, p. 85).

En primera instancia, se hace una distinción entre estados conscientes intencionales conscientes y no conscientes. No todo estado intencional es consciente y no todos los estados conscientes son intencionales. El criterio para considerar a un estado mental como tal sería que los estados neuronales que no son conscientes pueden ser comprendidos como mentales solamente con la condición de que sean capaces en principio de convertirse en estados conscientes, por ello Searle sostiene que “los estados mentales no conscientes deben de ser pensables conscientemente para ser un estado mental en lugar de ser un proceso cerebral no consciente” (Searle, 1983). Asimismo, hace una distinción entre estados cerebrales no conscientes como la secreción de determinados neurotransmisores, que son estados físicos del cerebro, pero que no implican una representación mental que sea ni cualitativa ni subjetiva como lo sería un estado consciente, a pesar de que la realización de estos estados cerebrales no conscientes pueden influir en la realización de los estados mentales conscientes, pero no en el nivel de realización que tendría la representación de una actitud proposicional. Nosotros somos totalmente inconscientes de la realidad que sucede en nuestro cerebro, por ello es no consciente. Sin embargo, sí hay ciertos estados no conscientes del cerebro que son capaces de causar fenómenos mentales conscientes (Searle, 1983). En la analogía que hace Searle sobre la computadora cuyas imágenes y palabras se guardan en el disco duro de otra forma, se expresa bien que los estados cerebrales no conscientes no necesariamente deben tener la misma forma que los conscientes, basta que puedan causarlos. Los estados mentales

no conscientes tienen una forma completamente diferente a los estados mentales conscientes, pero aun así, en tanto que tienen la capacidad de causarlos, se les considera estados mentales no conscientes.

Esta distinción que hace Searle no es compatible con otras teorías dentro de la ciencia cognitiva. Una de ellas es la de la gramática universal de Chomsky. Él sostendría que la facultad lingüística contiene un conocimiento innato de las reglas lingüísticas, que sería una sintaxis que subyace y posibilita la adquisición del lenguaje natural. Esto sería una suerte de estado inicial de la facultad lingüística. Durante el crecimiento, los seres humanos tienen experiencias del lenguaje que hacen que surja el conocimiento de lenguajes como el español o el inglés que comprenderíamos como la competencia lingüística. Es decir, que cuando aprendemos un lenguaje, para Chomsky, estamos siguiendo una serie de reglas no conscientes que constituyen la gramática universal (Chomsky, 1957). Esta teoría parecería bastante compatible con la noción del lenguaje del pensamiento de Fodor en tanto que, ambas normas que subyacen la competencia lingüística no son representables mediante un lenguaje natural, sino que obedecen a reglas sintácticas que posibilitarían que haya un contenido semántico en forma de lenguajes naturales, estas reglas serían de orden computacional.

El problema que Searle encuentra con esta postura es que si las normas que se están cumpliendo para que alguien aprenda un lenguaje por medio de una gramática universal que subyazca a las facultades lingüísticas es que las reglas son meramente sintácticas y sus procesos no serían el tipo de estados que podrían causar un estado mental consciente (J. Searle, 1999). Podríamos tratar de utilizar la metáfora de la computadora y los estados conscientes y no conscientes de la siguiente manera: La computadora no tiene que estar apagada para que las palabras y las imágenes que están en el monitor no se nos presenten, por ejemplo, cuando nosotros tenemos un archivo abierto podríamos pensar que es parecido a un estado mental consciente, pero hay muchos otros archivos que pueden estar cerrados, estos los podríamos asociar con los estados mentales no conscientes, incluso podríamos pensar en el tipo de archivos que el sistema operativo está ejecutando y que no somos conscientes de ellos. Si uno de estos archivos es borrado o llega a fallar, la computadora nos avisará que hay una falla en el sistema. De la misma manera, nosotros tenemos estados mentales no conscientes que explican causalmente nuestro comportamiento, estados mentales que no necesariamente puedan ser conscientes. Searle plantearía que esta

aproximación no sería coherente, porque no se establece en virtud de qué estos procesos son mentales, no explicarían que haya estados intencionales, o semánticos, sino meramente sintácticos. Este filósofo sostiene que los estados mentales genuinos deben de funcionar causalmente tanto cuando son consciente como cuando no lo son.

## La aproximación computacional digital como una explicación insuficiente de la mente

La teoría computacional de la mente trata de explicar ciertos poderes mentales como la inteligencia y la intencionalidad. En la aproximación mecanicista de Babbage lo que se trató de hacer fue replicar cierto tipo de trabajo que no solamente se cernía al tipo físico, como la distribución de funciones en la producción material, sino que se hizo un esfuerzo bastante loable para mecanizar procesos que, hasta ese momento, se comprendían como mentales. Turing trató de dar una definición de la inteligencia mediante el Test de Turing, y a la par desarrolló su famosa máquina teórica. Sin embargo, su definición no logró ser suficiente para describir los poderes inteligentes por su carácter conductista.

Lo que sí fue un éxito fue la máquina de Turing, la razón de ello fue que sentó las bases para explicar los procesos que podrían replicar algunas facultades que los humanos, mediante nuestros poderes mentales, podemos realizar. Como se explicó en el primer capítulo, esta máquina es teórica y esto supone que no es necesario tener un mecanismo que la soporte, uno la podría implementar solamente con una hoja y un lápiz, y ni siquiera eso, es implementable en nuestra mente. No importa en donde se ejecute un programa como la máquina de Turing, lo que importa es su naturaleza algorítmica.

Ahora bien, la máquina de Turing es muy importante en términos de la historia de la computación, porque representa un hito entre dos tipos de procesamiento computacional, lo que Piccinini clasifica como computación análoga, que tendría que ver con un tipo de procesamiento mecánico en el cual el programa está encarnado y tiene una relación de identidad con los engranes de la máquina, y la computación digital, que como bien lo dice su nombre, lo que procesa son dígitos (Piccinini, 2010). Ambos tipos de procesamientos computan símbolos, sin embargo, la diferencia radica en que la máquina de Turing es considerada multipropósito, sus procesamientos simbólicos no están limitados por las

relaciones mecánicas en su interior, de hecho, ni siquiera tiene interior. De ahí que la máquina de Turing de cabida al principio de realizabilidad múltiple.

La teoría computacional de la mente es un intento de dar una explicación de la mente mediante este principio, por lo que es una forma de funcionalismo. Como hemos visto, el Funcionalismo argumentaría que no importa el material donde se implemente un programa porque podemos ejecutar, por ejemplo, un procesador primitivo como una compuerta Y tanto con circuitos eléctricos como con cajas con gatos y ratones en su interior. El funcionalismo trata de comprender la mente mediante el análisis funcional, que es una forma de reducción científica, se tratan de explicar fenómenos complejos mediante sus componentes que son más simples. En el caso del Funcionalismo en la ciencia cognitiva lo que se trataría de abordar serían los procesadores primitivos que den cuenta de un fenómeno tan sofisticado como la mente. Para los funcionalistas los estados mentales, que podemos asociar con las actitudes proposicionales tales como creencias y deseos, se explican por medio de su constitución interna, pero solamente por su función dentro de un sistema cognitivo. La identidad de un estado mental estaría determinada por sus relaciones causales en un proceso que implicaría sus inputs, estados internos y outputs.

Por ello, apunta Block (1995b), que dentro de una explicación funcionalista la irrelevancia de la realización del hardware haría que dicha explicación sea incompatible con una teoría que explique la mente solamente en términos biológicos. En los términos de Copeland (1996b), la relación de modelaje entre un objeto y su diseño como un procesamiento computacional no sería biológico. Para los modelos computacionales no importa cómo se implementen los procesadores primitivos, lo que importa son las funciones que desempeñan. Dentro de la ciencia cognitiva se pretende que las explicaciones no sean incompatibles, sino que se complementen. El problema está en la asunción del principio funcionalista, ya que la teoría computacional no sería biológica, solamente importarían las equivalencias computacionales entre las funciones cerebrales y las funciones de la computadora, mientras que en la teoría biológica se comprende de manera distinta la naturaleza de la inteligencia humana.

Parecería que el punto clave que desestimaría la computación digital como una aproximación adecuada para explicar la mente tendría que ver con el hecho de que no podemos asumir que

el cerebro es un motor sintáctico que conduce un motor semántico, y que la conciencia como la característica esencial de la mente tiene un carácter irreductible.

En primer lugar, cuando se asume que el cerebro es un motor sintáctico que conduce un motor semántico lo que se acepta es que el contenido semántico, es decir, el contenido intencional, se da gracias a las estructuras sintácticas que operan los símbolos, y estos símbolos, en virtud de su relación isomórfica con sus referentes, como la que tendrían los numerales con los números, dotarían al sistema de función semántica. Sin embargo, esta aproximación que pretende dotar de intencionalidad a una estructura sintáctica no termina de dar cuenta de la intencionalidad como una característica de la mente, la razón de ello es que, a pesar de que sí podemos aceptar que dentro de gracias a una computadora, sí se pueden realizar ciertas operaciones matemáticas, esto solamente se da gracias a que nosotros sí comprendemos los significados de los números, nuestra representación no es digital. Al fin y al cabo, aquellos que realizan las operaciones matemáticas somos nosotros por medio de las computadoras que, gracias a sus estructuras sintácticas, resultan herramientas muy efectivas. Es innegable que las computadoras procesen símbolos, como los dígitos, pero también es bastante cuestionable que, de hecho, un sistema computacional comprenda el contenido intencional que portan dichos símbolos. La razón de ello es que el contenido intencional que se procesa dentro de una computadora no es original o primitivo, la relación causal adecuada para que haya intencionalidad primitiva no se da en la computadora, se da cuando una mente se relaciona con su ambiente. Los símbolos que están siendo procesados dentro de una computadora solamente tienen intencionalidad derivada de la misma manera que los termómetros y los indicadores de gasolina en los autos tienen contenido intencional.

Searle sostiene que podemos poner condiciones de satisfacción en aquello a lo que dotamos de intencionalidad derivada, como serían las preferencias o las palabras escritas, esto gracias a nuestra interpretación de ellos, pero su modo de intencionalidad no es independiente de nuestras interpretaciones. En su libro "Intentionality" (Searle, 1983) sostiene que los estados intencionales tienen condiciones de satisfacción. Por ejemplo, para una creencia, estas condiciones serían las que determinarían sus condiciones de verdad, para las experiencias sensoriales, las condiciones bajo las que caería mostrarían cuando éstas son verídicas. Estas condiciones de satisfacción permitirían que la mente pueda ser dirigida hacia los objetos en el mundo, y por supuesto, este tipo de relaciones tendrían que establecer el tipo de relación

causal de la mente con el mundo que explicarían la intencionalidad primitiva. Es decir que para Searle la clave para comprender la intencionalidad como representación son las condiciones de satisfacción. La noción de representación que utiliza Searle difiere de la se utilizaría en la teoría computacional de la mente porque, para él, una representación se define por su contenido y su modalidad, no por su estructura formal. Sostiene que, de hecho, no ve un sentido claro en el cual toda representación deba tener una estructura formal en el sentido que las oraciones tienen una estructura sintáctica formal.

Las condiciones de satisfacción son aquellas condiciones que, como determinadas por el contenido intencional, deben obtenerse si el estado es satisfecho. Por esta razón la especificación del contenido ya es una especificación de las condiciones de satisfacción. Por lo tanto, si yo tengo la creencia de que está lloviendo, el contenido de mi creencia es: que está lloviendo. Y las condiciones de satisfacción son: que está lloviendo –y no, por ejemplo, que el cuello está mojado o que hay agua cayendo del cielo. Dado que toda representación – ya esté hecha por la mente, el lenguaje, fotos o cualquier otra cosa. es siempre bajo ciertos aspectos y no otros, las condiciones de satisfacción son representadas bajo ciertos aspectos. (Searle, 1983)

El otro problema al que se enfrenta la teoría computacional de la mente es el de que sus representaciones, en tanto que están construidas sobre una estructura sintáctica que es meramente formal, pasan de largo cuando hablamos de que los estados mentales, para ser considerados como tales, son conscientes y, por lo tanto, son internos, cualitativos y subjetivos. Como se ha dicho anteriormente, es gracias a estas características que la conciencia resulta irreductible, especialmente por su carácter subjetivo.

Cuando hablamos de que la conciencia tiene este carácter nos referimos a un problema ontológico, es decir, que ciertos estados mentales tienen un modo de existencia de primera persona. Los estados conscientes solamente existen desde el punto de vista de un agente. Por eso es tan difícil para algunos filósofos que pretenden eliminarla el hacerlo, porque un objeto cuya existencia tiene una ontología de primera persona no puede ser reducida a una explicación con ontología de tercera persona. Esta es una de las características estructurales de la conciencia, y probablemente la que más resulte problemática por los principios del tipo de la explicación reduccionista que se pretende dar.



Además de esta característica estructural, Searle señala que la conciencia siempre tiene una forma unificada. La vida consciente de cualquier persona no solamente se refiere a un objeto, sino que tenemos una serie enorme de estímulos muy diversos cuando nos relacionamos con el mundo. Por ejemplo, en este momento estoy frente a una computadora, siento mi peso sobre la silla, siento el aire que entra por la ventana y roza mi piel, siento mis dedos haciendo presión sobre el teclado, etc. Sin embargo, todos estos estímulos no se me presentan de manera fragmentaria, sino que estoy teniendo una experiencia que los engloba a todos, esa experiencia está unificada. Searle enfatiza que la unidad de la conciencia, más que nada, refuta la visión de que hay distintos tipos de conciencia. De hecho, hay una distinción entre el pensamiento y la sensibilidad, sin embargo, ambas suceden al mismo tiempo en el mismo campo de conciencia. Los estados mentales conscientes tendrían un contenido distinto, pero ambos son parte de la misma experiencia consciente (Searle, 1983). Además, los estados conscientes siempre son estructurados. A menos que suframos de alguna enfermedad. El cerebro siempre le dará una estructura a los estímulos o inputs que entren su sistema de manera que sean coherentes. Nosotros estructuramos nuestras experiencias como un todo, por más que tengamos distintos inputs de nuestros entornos.

Como se ha mencionado anteriormente, la conciencia tiene una función esencial en nuestra supervivencia en el mundo, ésta nos da acceso al mundo. Es por ello que asumimos que la conciencia está unida a la intencionalidad. A pesar de que hay muchos estados intencionales y muchos otros estados conscientes que no son intencionales, hay una conexión entre la intencionalidad y la conciencia en este sentido: “La atribución de un estado mental a un agente es, ya sea una atribución de un estado consciente o la atribución de un estado del tipo de cosa que podría ser consciente” (Searle, 1983).

Otra característica de la conciencia es que todos los estados conscientes tienen un estado anímico, una noción muy parecida a lo que Heidegger llama “disposiciones afectivas” en la experiencia de su *Dasein* en la experiencia del mundo. Nosotros podemos estar angustiados, tristes, enojados, contentos o eufóricos, y este estado siempre es parte de la experiencia consciente que tenemos.

Asimismo, la conciencia tiene grados de atención, es cierto que, en los términos de Brentano, ser consciente es ser consciente de algo. Sin embargo, existen ciertos grados de atención. Searle expresa esto como una distinción entre el centro y la periferia creada por nuestros

variados grados de atención. Él utiliza la metáfora de una lámpara, cuando alumbramos algo en la oscuridad la enfoca de un lugar a otro, generalmente apuntando a las cosas que nos interesan más, hay otras cosas que quedan alumbradas, pero resultan periféricas. Además, otra característica relacionada con ésta es que existe una distinción entre el centro y la periferia creada por los grados de atención, los estados conscientes generalmente tienen un sentido de su situación. Esto es a lo que Searle llama las condiciones de borde. Cada estado consciente tiene un sentido de nuestra localización en el espacio y tiempo, a pesar de que ésta no sea un objeto de nuestra conciencia.

La conciencia siempre tiene grados de familiaridad, esto quiere decir que nuestras experiencias siempre se nos presentan como un continuo en un espectro que va desde lo más familiar hasta lo más extraño. Este aspecto de la familiaridad es muy difícil de romper, dado que los hechos de la intencionalidad, a saber, que toda representación mental se da desde un aspecto. Este aspecto se refiere al parecido que podemos encontrar entre las cosas con las cuales interactuamos, y desde él conocemos y reconocemos las cosas conceptualizándolas como casas o sillas, personas o autos, estas categorías son aspectos con los que estamos familiarizados. Y en gran medida esta sería una de las capacidades que explicarían que podemos identificar y reidentificar objetos.

La siguiente característica es que las experiencias conscientes se refieren a cosas más allá de ellas. Nosotros nunca tenemos una experiencia aislada, sino que siempre vienen acompañadas de otras experiencias. Esto tiene que ver con la característica interna de la conciencia, los estados conscientes están en relación con otros en una red de estados que se afectan causalmente los unos a los otros. Por ejemplo, cuando estamos recordando un evento familiar, nos es imposible dejar de pensar en todos los eventos que sucedieron en él, ya sea una conversación que a su vez nos refiere a otros eventos, o el chiste que contó alguien. En este sentido, parecería que la conciencia tiene una suerte de intertextualidad.

Y, finalmente, los estados conscientes siempre son placenteros o desagradables en algún grado. Esto tiene que ver con la característica del estado anímico y la familiaridad. Los estados conscientes siempre tienen un rango de distintos grados de placer o displacer, y por supuesto, la misma experiencia consciente puede contener rasgos placenteros o desagradables.

Estos aspectos estructurales que están íntimamente ligados a las características esenciales de la conciencia, a saber, que es interna, cualitativa y subjetiva, presentan muchos problemas para una explicación que reduzca los estados mentales a una estructura sintáctica formal. La irreductibilidad de la conciencia está sumamente ligada a estas características, sobre todo cuando su tipo de explicación trata de reducir objetos cuya existencia dependen de un agente, porque toda la noción de la conciencia, en primer lugar, existe para darle un nombre a los fenómenos subjetivos de la primera persona. A pesar de que la conciencia sea un fenómeno, que en tanto que tiene que ver con los estados cerebrales, es biológico, su ontología de primera persona la hace imposible de ser reducida a fenómenos objetivos de tercera persona. Si al funcionalismo y al cognitivismo les interesa dar una explicación de la mente, entonces dicha aproximación no sería suficiente para dar cuenta de los aspectos que la caracterizan como lo serían los estados mentales conscientes. Por supuesto que han hecho avances importantes, han desarrollado arquitecturas cognitivas como modelos computacionales de la mente que podrían explicar características muy importantes, por ejemplo, los procesos inteligentes. Pero es difícil aceptar que, de hecho, sus descripciones abarquen todo lo que implicaría el procesamiento mental de los estados conscientes. Se han desarrollado programas altamente sofisticados que logran emular muchos de los poderes mentales, sobre todo con respecto a procesos inteligentes. Pero distan mucho de explicar las características que acompañan a un estado mental, cuando estos se caracterizan por ser conscientes con todo lo que ello implica.

Turing desarrolla su máquina como un programa teórico que manipula símbolos en notación binaria. Pero como se explica en el argumento de la Habitación China, el programa, es decir, la estructura de procesamiento sintáctico es ciego a los significados de los símbolos que procesa, y por tanto no tiene comprensión del lenguaje. Los símbolos tienen un tipo de intencionalidad derivada, pero esa intencionalidad derivada no es independiente, al contrario, depende totalmente de que alguien interprete los símbolos, y ese alguien debe de tener conciencia.

## Las implicaciones ontológicas del funcionalismo y su escisión del mundo

El funcionalismo opera bajo el supuesto del principio de la realizabilidad múltiple, esto quiere decir que un programa computacional se puede ejecutar sin importar el material en donde se implemente. De ahí la consigna funcionalista de que lo relevante no es el hardware sino el programa que se está corriendo. Éste podría caracterizar los estados mentales como funciones dentro de un sistema, por ejemplo, un dolor sería causado por una herida y eso produciría la creencia de que algo le ha sucedido al cuerpo. Pero el dolor solamente sería reducido a la ejecución del programa del dolor. Por ello no importaría que este programa sea ejecutado en un material biológico, en silicio o mediante todos los habitantes de China. Lo que importaría serían las funciones que generan los patrones formales de activación de una red neuronal, circuitos eléctricos o la red telefónica de una nación. Por lo tanto, criaturas que no tengan cerebros como los nuestros podrían compartir sensaciones, creencias y deseos, sin importar el comportamiento ni su organización interna siempre y cuando se cumplan las condiciones formales para poder ejecutar el programa del dolor. Por ello, el funcionalismo sostendría que los estados mentales pueden ser realizados en múltiples materiales.

La discusión de Searle está directamente ligada a la noción de inteligencia artificial fuerte, que a su vez se vincula con la idea de que un programa computacional como SAM podría comprender historias.

Turing se hace la pregunta sobre la posibilidad de que las máquinas puedan pensar (1950), pero argumenta que ese planteamiento se puede reemplazar por la pregunta de si es teóricamente posible para una computadora digital de estados finitos que esté provista con una tabla de instrucciones, o un programa, dar respuestas a preguntas que podrían engañar a un interrogador para que piense que es un ser humano. Turing argumentaría que sí, gracias a que, si una máquina puede ser descrita como un cerebro, entonces lo único que hace falta es programar una computadora digital para imitar las funciones cerebrales y entonces así también será un cerebro. Es decir que para él, la inteligencia es un subproducto de la función computable de una máquina con independencia del material en donde esté implementada (González, 2011b). Turing identifica los pensamientos como estados de un sistema definido

solamente por sus roles en la producción de otros estados internos y outputs como las palabras, y esta perspectiva tiene mucho en común con las teorías funcionalistas contemporáneas.

En relación con este punto, González sostiene que el funcionalismo es lógicamente compatible con el dualismo, dada la irrelevancia del material en donde se implemente el programa, lo que importa es el programa mismo. Esto implicaría, más allá de un problema técnico sobre la implementación de los programas, una discusión entre dos problemas de tipo ontológico, es decir, la postura materialista y el dualismo.

El tema de la mente y la conciencia son problemas filosóficos bastante antiguos, y que determinan nuestra noción del mundo. Este problema ya ha sido mencionado, sobre todo cuando se plantean problemas como el hecho de que a partir de estados físicos no mentales puedan aparecer estados mentales conscientes.

Si piensas que la conciencia como un tipo de fenómeno separado y misterioso, distinto del material o de la realidad física, entonces parece que estás forzado a lo que tradicionalmente se le llama “dualismo”, la idea de que hay básicamente dos diferentes tipos de fenómenos o entidades en el universo. Pero si tratas de negar el dualismo y negar que la conciencia existe como algo irreductiblemente subjetivo, entonces parece que eres forzado al materialismo. Eres forzado a pensar que la conciencia, como la he descrito, y de hecho como la experimentamos, no existe realmente. Si eres un materialista, entonces eres forzado a decir que en realidad no hay tal cosa como una conciencia con una ontología subjetiva con una primera persona. (Searle, 1999).

Existen dos tipos de dualismo, el de sustancias y el de propiedades. De acuerdo con el primero, existen dos tipos de entidades que son radicalmente distintas y su tipo de existencia se excluye entre sí. Este tipo de dualismo es conocido como el dualismo cartesiano. El criterio de Descartes consiste tratar de distinguir las sustancias gracias a que se nos aparecen de manera clara y distinta. En sus *Meditaciones Metafísicas* dice que, si podemos concebir una cosa de manera clara y distinta como separada de la otra, entonces eso es suficiente para que tengamos certeza de que las dos cosas son distintas. Por ello postula la existencia de dos tipos de entidades, la *res cogitans* y la *res extensa*. La segunda, es decir, la cosa extensa, se refiere a los objetos que ocupan un lugar en el espacio, que tienen extensión, como lo serían nuestros

cuerpos. Y la primera, la cosa pensante, es aquella que puedo concebir sin la necesidad de aludir a algo físico con extensión. El punto del famoso *Cogito Ergo Sum*, es probar la existencia del pensamiento sin la necesidad de dar una explicación de mi identidad con mi cuerpo. Descartes asevera: “Tengo una idea clara y distinta de mí, en cuanto que soy simplemente una cosa no extensa pensante; y, por otra parte, tengo una distinta idea del cuerpo en cuanto esta es una cosa extensa no pensante. Y, en consecuencia, es cierto que soy muy distinto a mi cuerpo, y que puesto existir sin este.” (Descartes, 2006). La distinción que hace Descartes con respecto a los humanos y a los animales sobre el entendimiento del lenguaje está íntimamente ligada a que, para él, la mente, en tanto que es *res cogitans*, tendría un carácter indeterminado y flexible. Mientras que los animales están determinados causalmente por lo que él llama “espíritus animales” que no deben ser comprendidos como *res cogitans*, más bien son ciertas disposiciones físicas en los animales que no permiten que tengan la flexibilidad de la conciencia de la cuales las mentes humanas sí gozan.

El otro tipo de dualismo es el de propiedades, en este lo que se sostiene es que hay dos tipos de propiedades de los objetos que son distintas en un sentido metafísico. Por ejemplo, las cosas tienen peso o temperatura. Este tipo de propiedades no son iguales al dolor o a un estado mental. Por una parte, tenemos ciertas descripciones de objetos, como las que haría la ciencia, que especifican propiedades en términos meramente causales. Sin embargo, parecería que necesitamos más que este lenguaje para describir un dolor. El dualismo de propiedades ocurre cuando la ontología de la física o de la química no son suficientes para explicar fenómenos como los mentales. Hasta donde hemos visto, parecería que Searle de alguna manera comparte esta visión del dualismo de propiedades, sin embargo, argumenta que no, tema será abordado en el siguiente apartado.

El problema con el dualismo es que nos obliga a pensar en un mundo dividido ya sea en sustancias o en tipos de propiedades y no nos permite comprenderlo unitariamente. Ahora bien, una postura que se muestra bastante convincente es la del materialismo, sobre todo cuando se trata de tener una explicación unitaria del mundo. En este punto de la historia contamos con la suficiente evidencia como para no aceptar los hechos de la ciencia moderna. Hasta donde sabemos, el mundo se explica desde la teoría atómica de la materia y la teoría evolucionista biológica (Searle, 1999). La base de estas teorías consistiría en que el universo consiste enteramente de materia, y en específico, de ciertas entidades que llamamos

“partículas” que están en campos de fuerza, y cuyos límites hacen que un sistema esté determinado por sus relaciones causales.

La ciencia cognitiva adhiere al materialismo, es difícil, hoy en día encontrar incluso algún filósofo que no acepte algún tipo de materialismo. El problema que se le presenta a éste es que un tipo de sustancia, como la pensante, funcione fuera de los límites de las estructuras causales que implicaría aceptar que todo es material. Para un materialista es inconcebible que la conciencia sea una entidad que no esté sujeta a las leyes naturales. Por ejemplo, el conductismo afirma que la mente se puede reducir a comportamiento y a disposiciones de comportamiento. El fisicalismo dice que los estados mentales son estados cerebrales.

El funcionalismo, y por lo tanto también la teoría computacional de la mente, también es compatible con el materialismo, pero solamente en el nivel de la implementación, ya que para esta perspectiva lo que es relevante para modelar un software que logre replicar ciertos procesos cognitivos que se dan en el cerebro no es el sustento material. Para el funcionalismo los estados mentales están definidos por sus relaciones causales. De acuerdo con los funcionalistas, cualquier estado de un sistema físico, ya sea un cerebro o una tarjeta madre, que tenga las relaciones causales entre sus inputs, sus estados internos y outputs, ya podrían contar como estados mentales. Asimismo, la teoría computacional de la mente dice que las mentes solamente son programas computacionales implementados en cerebros, y que, si logramos descifrar todas sus funciones, entonces podemos replicar mentes.

El punto de González cuando afirma que el funcionalismo es compatible con el dualismo porque:

[...]si bien el funcionalismo y el principio de realizabilidad múltiple no implican de manera obvia al dualismo, la compatibilidad de ambos con esta concepción y con la idea de que el material no importa genera problemas metafísicos, toda vez que el funcionalismo y otras teorías en principio *fisicistas* son justamente planteadas como alternativas materialistas a la propuesta de Descartes. (González, 2011b)

Es decir, a pesar de que Turing trata de dar una respuesta materialista, y ésta da lugar a un tipo de funcionalismo que parece compatible con el dualismo cartesiano. Sin embargo, gracias al principio de realizabilidad múltiple parece que ya está haciendo una distinción entre la implementación material del programa y el programa mismo. Es cierto que la

capacidad de la Máquina de Turing puede simular cualquier sistema algorítmicamente calculable, y el proyecto de la inteligencia artificial clásica o fuerte contempla que las computadoras puedan realizar tareas que solamente se les adjudicaban a los poderes mentales humanos. Sin embargo, como hemos visto, existen una serie de argumentos que no permitirían que la inteligencia artificial fuerte sea suficiente para emular una mente consciente plenamente, de hecho, su rango de acción sería bastante limitado. Tal como sostiene González, “La IA es una extraña mezcla de Conductismo y Dualismo. Es conductista en su aceptación del Test de Turing, pero en un nivel más profundamente filosófico es dualista, porque rechaza la idea de que la conciencia y la intencionalidad son fenómenos biológicos ordinarios como la digestión.” (González, 2011b).

El problema que tiene el funcionalismo, en principio, es que hace una escisión entre la implementación material del programa y el programa mismo. El programa con su carácter algorítmico es pensado solamente como un diseño formal permitiendo que haya una relación de modelaje que permite que algo sea un sistema computacional sin importar la materia en la cual esté implementado.

Si asumimos que para que los estados mentales sean considerados como tales, entonces tenemos que aceptar que su característica esencial es la conciencia. Pero la conciencia tiene un carácter irreductible. Parecería que estamos en una encrucijada entre hacer una reducción materialista de la mente y tomar una postura eliminativista de la conciencia, o aceptar un dualismo ya sea de propiedades o de sustancias. Pero parecería, también, que tenemos otra opción: No comprender los procesos mentales como meros estados físicos que cuya esencia se reduzca a estructuras sintácticas, sino aceptar que la materia sí importa y entonces, comprender la mente bajo una perspectiva biológica que pueda dar cabida a sus características internas, cualitativas y subjetivas, en un nivel de realización distinto al de la sintaxis.

### La realización biológica de la mente dentro de una ontología monista

El problema del materialismo y del dualismo son problemas filosóficos que parecen irresolubles cuando tratamos de dar cuenta de la naturaleza de los estados mentales conscientes. Parece que siempre tenemos dos alternativas, de las cuales ninguna puede ser



aceptable ya sea por comprometernos con objetos cuya existencia no sea explicable por medio de nuestro conocimiento y la evidencia que tenemos para poder justificarlo. Muy probablemente estaríamos multiplicando las entidades innecesariamente. Por otro lado, parece que tenemos el problema de aceptar que nuestras experiencias internas, cualitativas y subjetivas no existen y que solamente son ilusiones. Pero las ilusiones solamente se le presentan a la conciencia. Parecería que nos enfrentamos a tener que tomar una postura y que hacerlo será crítico.

En el caso de la conciencia y del problema entre la distinción entre la mente y el cuerpo, se nos ha dicho que tenemos que elegir entre el dualismo, que insisten en la irreductibilidad de lo mental, y el materialismo, que insiste que la conciencia debe ser reductible y por lo tanto eliminable a favor de una explicación de la mente puramente fisicalista o material (Searle, 1999).

Este filósofo, tiene un método para resolver este tipo de problemas, lo que él llama el choque de posiciones default. Estas pueden ser comprendidas como los enunciados que sostenemos cuando militamos en determinadas posturas ontológicas. Las posiciones default son las opiniones pre reflexivas de tal manera que cualquier salida de ellas requiere un esfuerzo consciente y un argumento convincente. No todas las posiciones default implican verdad, de hecho, los grandes hitos de la historia implican un cambio de las posiciones default, eventos como la llegada de los europeos a América cambió toda la manera de concebir el mundo tanto para europeos como para americanos. La física newtoniana logró matematizar los comportamientos de la naturaleza y expresarlos mediante ecuaciones, lo cual cambió en gran medida la manera en que conocemos como conocimiento, que es el principio de lo que Kant llamaría el giro copernicano. Las características que se presentaron en el apartado anterior son nociones centrales que determinarían las posiciones default tanto del materialismo como del dualismo.

Para comprender la postura de Searle, tenemos que saber cuáles son las posiciones default que acepta, las cuales enumera de esta manera:

- Hay un mundo real que existe independientemente de nosotros, independientemente de nuestras experiencias, nuestros pensamientos y nuestro lenguaje.
- Tenemos un acceso perceptual directo al mundo a través de nuestros sentidos, especialmente el tacto y la vista.
- Las palabras en nuestro lenguaje, palabras como conejo o árbol, típicamente tienen significados razonablemente claros. Gracias a sus significados, pueden ser utilizados para referir y hablar sobre objetos reales en el mundo.
- Nuestras afirmaciones son generalmente verdaderas o falsas dependiendo de si corresponden a cómo son las cosas, esto es, a los hechos en el mundo.
- La causación es una relación real entre objetos y eventos en el mundo, la relación por la cual un fenómeno, la causa, causa otro, el efecto. (J. Searle, 1999)

Searle apunta que las implicaciones de asumir el materialismo en su totalidad son peligrosas porque se termina eliminando la conciencia con su ontología de primera persona, y, por otra parte, si asumimos el dualismo, entonces no podemos conocer la mente en tanto que sería una sustancia ajena a nuestro campo. Él cree que hay una realidad externa a nosotros y tenemos una relación consciente con ella con ella. Y como se ha dicho antes, la herramienta evolutiva que tenemos para lograr es y sobrevivir es la conciencia con su carácter intencional. Sin embargo, la diferencia que tendría este filósofo con las otras posturas materialistas no estaría exactamente en que difiera con las posiciones default, sino en el tipo de explicación que se tiene que dar de la conciencia, y por ende de la mente. Por ello Searle toma una postura que llama “naturalismo biológico”, esto quiere decir que no deja de aceptar que hay una realidad independiente de nosotros, pero a la vez tiene que explicar que existan fenómenos mentales con una ontología de primera persona sin la trampa de caer en el dualismo, por ello la parte biológica de su realismo radicaría en el tipo de explicación de la mente que él propone que debiera de haber, y que en este caso apelaría a una realización biológica de los estados mentales conscientes.

No es posible desestimar los avances explicativos de la ciencia cognitiva con respecto a ciertas facultades de la mente, sin embargo, la aproximación computacional digital parece ser insuficiente. La razón de ello está en la distinción que hemos visto entre la intencionalidad primitiva y la derivada. Las adscripciones de intencionalidad que se le atribuyen a los

sistemas computacionales por medio de la noción del motor sintáctico que conduce un motor semántico son metafóricas. Decir que una entidad tiene intencionalidad es lo mismo que decir que se comporta como si tuviera intencionalidad, cuando en realidad no es así. Toda intencionalidad derivada es derivada de la primitiva. Es importante hacer especial énfasis en esto porque muchos autores han tratado la intencionalidad derivada como si esta pudiera explicar la intencionalidad primitiva. Searle sostiene que la conciencia y la intencionalidad, aunque son características de la mente, son independientes del observador en el sentido de que, si somos conscientes o tenemos estados intencionales, la existencia de estas características no depende de lo que alguien fuera de mí piense. La postura de Searle es que la conciencia es un fenómeno natural, y la intencionalidad también lo es. Es decir, la mente en tanto que es parte del mundo también responde a las leyes naturales.

La intencionalidad tiene poderes causales, por ejemplo, cuando tenemos el estado consciente de tener sed, entonces ese estado causa que nos dispongamos a ir por un vaso de agua. Esto se da gracias a un proceso neurobiológico. Este es el tipo de explicaciones que, sostiene Searle debemos de dar sobre la mente. Para él no podemos divorciar, por ejemplo, la función digestiva del aparato digestivo. De la misma manera, no podemos escindir el pensamiento del sistema nervioso humano. Esto se podría explicar por medio de los niveles de realización de la mente. En un nivel muy bajo lo que nos encontramos es un entramado de relaciones físicas y químicas que dan lugar a procesos biológicos como la secreción de neurotransmisores o de la sinapsis en las neuronas. Pero no es sino la conjunción de todos estos elementos que pueden hacer posibles funciones biológicas como la digestión, respiración o el pensamiento.

Una postura que parece cumplir con las condiciones biológicas que exigiría una postura como la de Searle, sin caer en una explicación que meramente apele a las relaciones sintácticas dentro de un sistema meramente formal, y sin dejar de lado las características biológicas de la mente sería la de Gualtiero Piccinini. A la vez, tampoco echa por la borda las nociones computacionales que permiten explicar algunas características de la mente. Él sostiene que todo el sistema nervioso es un sistema computacional (Piccinini, 2010).

Para este la ciencia cognitiva asimila la explicación mecanicista a la explicación computacional, es decir que asimila los procesos mecánicos a cómputos. Explica que un proceso mecanicista es una manifestación de un conjunto de operaciones organizadas de

cierto modo, por lo que un proceso mecanicista es un cómputo y una explicación mecanicista es computacional. Algo así está detrás de la visión de que todo es computacional, y que algún formalismo computacional como las máquinas de Turing captura todo lo que puede hacerse mecánicamente. Por ello hace una distinción entre la explicación mecanicista y el análisis funcional. La primera tiene que ver con los componentes concretos de las máquinas, mientras que la segunda es mucho más abstracta en tanto que solo ve funciones (Piccinini, 2016b).

Además, sostiene que la explicación computacional debe de ser mucho más que la simple apelación a inputs, estados internos y procesos. Ya que de otra manera, se puede perder la especificidad de cierta clase de mecanismos y trivializar la teoría computacional de la mente. Es decir que, se necesitan ciertas restricciones a la naturaleza de los inputs, outputs, estados internos y procesos. Que es el mismo problema por el que se preocupa Copeland cuando tiene que establecer una relación de modelaje que explique la computación (Copeland, 1996b). Lo mismo pasa con la explicación mecanicista, su fuerza está en sus referentes concretos. Si esto se abstrae bajo una noción de computación, entonces todas las explicaciones se ven parecidas y todo caería bajo la lupa de los estados internos. Si una explicación mecanicista fuera lo mismo que una explicación computacional, entonces todo artefacto y órgano biológico sería un organismo computacional, y en realidad nunca se les ve de esta manera. Si fuera así, el cerebro sería tan computacional como el estómago. Piccinini hace una distinción importante entre el proceso mecánico y el computacional. La computación merece ser explicada mecánicamente, pero no debe de ser confundida con todo mecanismo. La máquina de Turing es la formalización de un mecanismo, sin embargo, como se mencionó en el primer capítulo, una estructura formal que no está implementada y que funciona en el nivel teórico no obedece a la causalidad, más bien, funciona gracias a la transitividad lógica de la misma manera que las fórmulas lógicas. Sin embargo, esto no quiere decir que no sea un tipo de explicación mecanicista, la computación digital y su aproximación a la mente, asumiéndola como el software del cerebro, pretende explicar los procesos cognitivos mecánicamente, pero asume que se puede implementar en cualquier material. Por su lado, un mecanismo ya implementado siempre tendría que tomar en cuenta la causalidad y los materiales en donde se es realizado.

El problema que tiene el análisis funcional es que piensa la aproximación computacional como una estrategia explicativa distinta y autónoma de la explicación mecanicista. Es decir

que, no es posible el pensar el programa con autonomía del material donde esté siendo implementado. En el funcionalismo los sistemas computacionales se explican por medio de la descomposición recursiva hasta llegar a los procesadores primitivos, y luego se explican las computaciones realizadas por los sistemas en términos de funciones.

La noción de computación que le debemos a Turing puede ser formulada en términos de cadenas de símbolos (dígitos, letras o palabras). Estos símbolos pertenecen a caracteres finitos, que son distinguibles por los mecanismos que los manipulan. Los sistemas computacionales digitales concretos son mecanismos cuya función es manipular cadenas de dígitos de acuerdo con reglas que generan una sintaxis de la cual dependen las cadenas de inputs para su procesamiento (Piccinini, 2010). Es decir, que una explicación computacional es entonces una explicación mecanicista en la cual los inputs, outputs y sus estados internos son cadenas de dígitos, y el procesamiento de cadenas puede ser capturado precisamente por las reglas sintácticas. En este sentido, la explicación computacional digital no se aplica a los seres humanos dado que nuestro contenido semántico no puede ser reducido a dígitos o a meras estructuras sintácticas, sino que para nosotros se requiere otro tipo de formato que pueda dar cuenta de los estados conscientes internos, cualitativos y subjetivos.

La propuesta central de Piccinini apela a lo que llama “mecanismos neuronales” y sostiene que puede haber una explicación mecanicista en la neurociencia.

Las funciones del cerebro (y generalmente aquellas del sistema nervioso) pueden ser descritas aproximadamente como el *feedback* de control del organismo y sus partes. En otras palabras, el cerebro está a cargo de producir un rango amplio de actividades realizadas por el organismo de tal manera que sea sensible al estado tanto del organismo y de su ambiente (Piccinini, 2010)

Lo que el sistema nervioso haría sería recolectar y llevar información para ser propiamente sensitivo. Ahora bien, el llevar información significa poseer transportadores que se correlacionan con el formato de la información que portan. La explicación mecanicista del sistema nervioso requiere apelar a estados internos que se relacionan con el resto del cuerpo y el ambiente de maneras apropiadas. Y cuando un sistema tiene este tipo de transporte de información y una función de procesamiento, decimos que ese sistema posee representaciones internas.

Piccinini enfatiza que estos vehículos del cerebro son picos generados por las neuronas, y estos se organizan en secuencias que se conocen como trenes de picos. Sus propiedades varían de neurona a neurona y son insuficientes para producir un efecto funcionalmente relevante. La explicación mecanicista en la neurociencia consistiría en especificar como es que estas organizaciones de trenes de picos constituyen las capacidades del cerebro y esto incluye sus capacidades cognitivas. Estas cadenas de picos no cuentan de ninguna manera como dígitos, la computación neuronal no es ni digital ni análoga. Por ello hace falta un gran camino que recorrer en la neurociencia. La computación digital no es suficiente para explicar los poderes sensitivos que implicaría la mente porque es ciega a los contenidos semánticos que conlleva la comprensión del lenguaje. Este tipo de contenidos tendrían que ser comprendidos como estados mentales conscientes, o al menos con la posibilidad de serlo. Esto supondría que como cualquier objeto consciente comprendería que tienen un tipo de ontología de primera persona, que son cualitativos y son internos. Tendríamos que hacer una aproximación que no desestime los logros de la ciencia cognitiva, pero que a la vez no elimine el carácter fenoménico que tiene la mente. Se debe de buscar un tipo de explicación más integral que no reduzca la conciencia a ilusiones y que no le quite su papel central en la comprensión de los estados mentales. Por ello, parecería mucho más conveniente no dejar de lado las explicaciones computacionales, así como no caer en realizar definiciones triviales de la computación, y aceptar que la definición de la computación neuronal no es trivial y que, de hecho, es una aproximación que cumple con la condición de tomar en cuenta el material biológico sobre el que se realiza la mente, toma en cuenta la relación de modelaje que permitiría que algo sea una computadora, y por ello, podría dar cuenta de los estados mentales conscientes con su carácter fenoménico gracias a la sensibilidad que implicaría el material biológico sobre el cual está construida. Es decir que el factor biológico no trivializaría la computación y también nos obligaría a pensar que el cerebro es una computadora sui generis.

## Conclusiones

La aproximación computacional digital de la mente no es suficiente para comprender nuestros estados mentales conscientes. La razón de esto es que el lente por medio del cual se trata de comprender la mente está basado en el funcionalismo y su escisión de los materiales

en los cuales los programas son implementados. Existen varias capacidades que en su momento fueron adjudicadas meramente al trabajo mental, sin embargo, sí es posible replicar este tipo de procesamientos cognitivos como la resolución de problemas matemáticos, jugar ajedrez, reconocimiento de patrones de escritura, etc., por medio de programas computacionales. Sin embargo, cuando se hace una aproximación computacional digital nos enfrentamos al problema de dar cuenta del contenido semántico que es procesado dentro del sistema. La visión funcionalista trataría de explicarlo aludiendo a la noción del motor sintáctico que conduce un motor semántico, de esta manera lo que se pretende hacer es reducir los estados mentales y sus significados a funciones dentro de un programa computacional. Se han hecho propuestas como el lenguaje del pensamiento de Fodor (1979), en donde se plantea que tenemos un sistema sintáctico que operaría las representaciones que en forma de outputs tendrían la misma forma que un lenguaje natural. Pero su estructura sería meramente formal.

El problema al que nos enfrentamos con este tipo de aproximaciones es que, como bien apuntaría el experimento de la Habitación China, los sistemas computacionales son ciegos a al contenido semántico de los símbolos procesados. El sistema no tiene comprensión de lenguaje. Lo que un programa computacional hace es manipular símbolos de manera tal que solamente por la estructura formal de los símbolos se conserve el contenido intencional que se les ha atribuido, esto quiere decir que una computadora no tiene la capacidad para traducir, solamente conservaría las estructuras sintácticas de un lenguaje por medio de la relación de modelaje sintáctica del software y la sintaxis de dos o más lenguajes naturales. Sin embargo, la computadora no tendría el tipo de historia causal necesaria como para comprender el significado de los símbolos procesados porque no tiene el acceso adecuado a los referentes. La noción de inteligencia artificial fuerte sostiene que ciertos programas computacionales pueden comprender historias, pero Searle demuestra que, si no hay comprensión lingüística, entonces las computadoras no podrían tener esa función.

Lo que se tiene que hacer en primera instancia es entender el tipo de intencionalidad que sería capaz de generar el tipo de relaciones causales apropiadas para que un estado mental cuente con la intencionalidad primitiva. Por ello Searle sugiere que la clave está en que debe haber una historia causal entre un cerebro y los objetos en su ambiente con los cuales interactúa, este tipo de relación necesita satisfacer determinadas condiciones de verdad que harían que

nuestras representaciones mentales tengan referentes y que sean verídicas (Searle, 1983). De otra manera, tendríamos sistemas sintácticos sin funciones semánticas como el experimento mental del Hombre del Pantano.

Pero ese no es el único problema con la aproximación de la computación digital. Searle afirma que una característica definitiva que determinaría a un estado mental como tal es la conciencia. Él plantea la conciencia como una función evolutiva que nos ha permitido relacionarnos con nuestro ambiente, así como con otros seres humanos por medio del lenguaje. Sin ella no podríamos sobrevivir en este mundo (Searle, 1999).

La conciencia tiene tres características esenciales: es interna, cualitativa y subjetiva. Y en esos términos es como experimentamos a un nivel fenoménico nuestras representaciones. Los estados mentales conscientes son internos en tanto que se realizan dentro de nosotros, específicamente en nuestros cerebros, y son internos ya que están en relación con un entramado de otros estados mentales. Tienen un carácter cualitativo que nos permite experimentarlos fenoménicamente, tienen una cierta manera de sentirse, desde realizar una operación matemática hasta gozar de una ópera de Verdi. Además, los estados internos tienen un tipo de ontología de primera persona, esto gracias a su carácter subjetivo. Un estado mental consciente solamente existe gracias a la relación que tiene un sujeto con su ambiente, y el sujeto tiene la autoridad de la primera persona a esas experiencias, cuando nos relacionamos con nuestro ambiente lo experimentamos en primera persona y esta experiencia solo es subjetiva. Por ello es que la conciencia tiene un carácter irreductible, porque los estados mentales solamente se dan en estos términos, su forma de existencia depende enteramente de la relación de los sujetos en una relación de sintiencia con su ambiente que se experimenta subjetivamente.

Todo estado mental consciente está relacionado con el mundo por medio de la intencionalidad primitiva que permite realizar representaciones en los términos de las características esenciales de la conciencia. Y gracias al carácter irreductible que tienen, no pueden ser entendidos como símbolos que son procesados por una estructura meramente sintáctica. Existen filósofos eliminativistas que tratan de reducir la conciencia de tal manera que se termina comprendiendo como una ilusión o, en el mejor de los casos, como un residuo de los procesos computacionales que se realizan en el cerebro. Sin embargo, si la conciencia es una ilusión, entonces no se le puede presentar a otra cosa que no sea ella misma. Para



Searle esta es una visión incorrecta porque él comprende a la conciencia como la característica más primordial de lo que entendemos por mente.

No todos los estados cerebrales son mentales, por ejemplo, existen estados internos del cerebro que consisten en la liberación de neurotransmisores, estos estados cerebrales no pueden convertirse en estados mentales. Sin embargo, lo que nos interesa en términos de estados que puedan tener intencionalidad son los estados cerebrales que puedan contar como estados mentales. Asimismo, no todos los estados mentales son conscientes, existen estados mentales no conscientes, pero les seguimos atribuyendo un carácter mental porque tienen la capacidad de causar estados mentales conscientes. Uno puede tener la creencia de que Marte es el cuarto planeta del sistema solar, pero cuando vamos a dormir esa creencia se transforma en otro tipo de estado mental no consciente, pero no deja de ser mental porque cuando despertamos y nos preguntamos por el cuarto planeta del sistema solar podemos traer esa creencia en forma de un estado mental consciente.

Para que tengamos un criterio adecuado para determinar cuándo un estado mental lo es, no podemos negar nuestras experiencias sensibles, y se debe de dar una explicación completa sobre ellos. Además, como bien apunta Copeland, esta explicación debe de poder ser sometida a una revisión empírica (Copeland, 1996b). Por ello no podemos inclinarnos por el dualismo, el cual nos impide tener acceso a los estados mentales como si fueran sustancias o propiedades que pertenecen a otro reino que no sea material. Debemos de comprender que la intencionalidad y la conciencia atienden a las leyes físicas y son parte del mundo natural.

Es cierto que la ciencia cognitiva ha hecho aproximaciones muy poderosas sobre la mente, no podemos desestimar sus avances en términos de arquitecturas cognitivas, pero tampoco podemos reducir los estados mentales conscientes a meras estructuras sintácticas. Por ello, parecería adecuado darle un giro a la aproximación computacional y abandonar la visión de la computación digital y comenzar a pensar la mente en términos biológicos. La teoría computacional de la mente se genera en términos mecanicistas, por ello es una teoría mecanicista. Sin embargo, no toda teoría mecanicista es una teoría computacional digital, como lo vimos al principio de la tesis, la máquina diferencial de Babbage se caracteriza por la identidad que tienen sus engranajes con la función que realiza. Pero la computación digital que tiene un fundamento funcionalista, a pesar de que no deja de ser mecanicista, escinde el programa del material en donde se implementa. Si queremos tomar en serio el problema de

la conciencia no podemos negar sus características, y sobre todo hay que tomar en cuenta que, dado el carácter irreductible de estados mentales conscientes como el dolor, es un criterio esencial para determinar si un estado mental realmente contaría como tal en tanto que tiene un modo de existencia subjetivo, interno y cualitativo. Sin embargo, si tenemos una explicación biológica, entonces el factor de la implementación material de un programa sí es relevante porque es el material biológico que permite que la conciencia en términos de sensibilidad. Incluso Dennett acepta que, si se llega a desarrollar un programa para el dolor, sería muy cuestionable en términos empíricos que se realice en otra cosa que no sea un cerebro orgánico (Dennett, 1993). Pensar a la mente como el software del cerebro es una idea bastante seductora, sin embargo, no deja de haber una escisión entre el software y el hardware. Pero si pensamos que los estados mentales se realizan como estados conscientes, con todo lo que ello implica, tal vez la mejor opción sería dejar el principio de realizabilidad múltiple y pensar que las características esenciales de la conciencia se dan en términos sensibles y la única manera de explicarlos sería por medio de otro tipo de explicación mecanicista, no una como la computación digital, pero sí una como la computación neuronal. La aproximación computacional comprende que hay una relación de modelaje que implica inputs, estados internos y outputs, este tipo de aproximación no parecería inadecuada para la comprensión de la cognición humana, tenemos inputs que entran por nuestros sentidos, son procesados por medio de estados cerebrales, de los cuales algunos son considerados mentales, y outputs como nuestros comportamientos. Pero el tipo de formato en el que se presentan estos inputs, estados internos y outputs difiere mucho al de la computación digital, además de que para procesar esos formatos se requieren determinados tipos de estados internos muy distintos a los de una relación meramente sintáctica, tendrían que ser relaciones neuronales cuyo contenido no sea digital.

Los seres humanos tenemos una capacidad para relacionarnos con nuestro ambiente que se da gracias a la intencionalidad, sin ella no podríamos si quiera sobrevivir, nuestra gran capacidad poética de representaciones se la debemos en gran parte a nuestra sensibilidad. Nuestras representaciones se nos presentan siempre de manera cualitativa porque se sienten de ciertas maneras, desde la realización de un problema matemático hasta la contemplación de un amanecer, son internas porque están dentro de nosotros en forma de estados cerebrales, pero también porque están en relación con un entramado de otras representaciones que

generamos durante nuestra vida consciente, y, sobre todo, son subjetivas, esto porque solamente existen para nosotros. Es innegable que estas características se dan gracias a nuestra sensibilidad y nuestra experiencia del mundo. Somos animales complejos con una capacidad de representar y de resolver problemas gracias a ello, y estas representaciones se presentan como estados mentales conscientes. La teoría computacional de la mente es una herramienta muy eficaz para poder comprender algunas de nuestras capacidades cognitivas, pero no todas y no las que más nos importan por su carácter fenoménico. Es posible dar cuenta de ciertas capacidades inteligentes por medio de su emulación en las arquitecturas cognitivas que ven a la mente desde el punto de vista de un programador. Pero la capacidad de representarnos el mundo conscientemente aún es un reto. Sin duda, la ciencia cognitiva ha hecho avances importantes en la comprensión de varios de estos procesos, pero el problema que se le presenta es que su explicación de la intencionalidad no toma en cuenta las características de la conciencia y puede caer en el eliminativismo. Su aproximación apela al isomorfismo del motor sintáctico conduciendo un motor semántico, pero los símbolos que son procesados son otorgados por mentes conscientes que utilizan estos procesos como herramientas. Por ello, debemos de poner a la computación digital en su justo lugar y aceptar el papel central de la conciencia y de la intencionalidad en la cognición humana, pero no desde una perspectiva digital, sino aceptar lo que ya sabemos, que somos seres biológicos que obedecemos a las leyes naturales y que gracias a nuestra inteligencia e intencionalidad, hemos podido representar el mundo, sobrevivir y desarrollar cultura.

## Bibliografía

- Block, N. (1978). Troubles with functionalism. En W. Savage (Ed.), *Perception and Cognition* (pp. 9--261). University of Minnesota Press.
- Block, N. (1995a). The mind as the software of the brain. En D. N. Osherson, L. Gleitman, S. M. Kosslyn, S. Smith, & S. Sternberg (Eds.), *An Invitation to Cognitive Science, Second Edition, Volume 3* (pp. 377-425). MIT Press.
- Block, N. (1995b). The mind as the software of the brain. En *An invitation to cognitive science* (2a. Edición). The MIT Press.
- Brentano, F. (2015). *Psychology from an empirical standpoint*. Routledge.
- Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.
- Chomsky, N. (1957). *Syntactic Structures* (Vol. 31, Número 2, pp. 245-251). Mouton.
- Churchland, P. S. (1983). CONSCIOUSNESS: THE TRANSMUTATION OF A CONCEPT. *Pacific Philosophical Quarterly*, 64(1), 80-95.  
<https://doi.org/10.1111/j.1468-0114.1983.tb00186.x>
- Cole, D. (2020). The Chinese Room Argument. En *The Stanford Encyclopedia of Philosophy* (Spring 2020). <https://plato.stanford.edu/archives/spr2020/entries/chinese-room>
- Copeland, B. J. (1996a). What is computation? *Synthese*, 108(3), 335-359.  
<https://doi.org/10.1007/BF00413693>
- Copeland, B. J. (1996b). What is computation? *Synthese*, 108(3), 335-359.  
<https://doi.org/10.1007/BF00413693>
- Cummins, R. (1975). Functional Analysis. *The Journal of Philosophy*, 72(20), 741.  
<https://doi.org/10.2307/2024640>
- Dennett, Daniel C. (1987). Fast thinking. En *The Intentional Stance*. MIT Press.
- Dennett, Daniel C. (1992). Quining *qualia*. En A. J. Marcel & E. Bisiach (Eds.), *Consciousness in Contemporary Science* (pp. 42-77). Oxford University Press.  
<https://doi.org/10.1093/acprof:oso/9780198522379.003.0003>
- Dennett, Daniel Clement. (1993). *Consciousness explained*. Penguin.
- Descartes, R. (2006). *Discurso del método ; Meditaciones metafísicas*. Espasa.
- Fodor, J. (1965). Explanations y Psychology. *Philosophy in America*.

- Fodor, J. A. (1979). *The Language of thought* (1st paperback printing). Harvard Univ. Press.
- Frankish, K., & Ramsey, W. M. (Eds.). (2012). *The Cambridge handbook of cognitive science*. Cambridge University Press.
- Frege, G., Mosterin, J., & Moulines, C. U. (1985). *Estudios sobre semántica*. Orbis.
- González, R. (2007). *The Chinese room revisited: Artificial intelligence and the nature of mind*. Katholieke Universiteit Leuven.
- González, R. (2011a). Máquinas sin engranajes y cuerpos sin mentes. ¿cuán dualista es el funcionalismo de máquina de Turing? *Revista de Filosofía*, 67, 183-200.
- González, R. (2015). ¿Importa la determinación del sexo en el Test de Turing? *Revista de Filosofía Aurora*.
- González, R. (2016). El entendimiento lingüístico en la Inteligencia Artificial: Una relación ambivalente con Descartes. *Revista eletrônica de investigação filosófica, científica e tecnológica*.
- Haugeland, J. (2003). Syntax, semantics, physics. En J. M. Preston & M. A. Bishop (Eds.), *Views Into the Chinese Room: New Essays on Searle and Artificial Intelligence*. Oxford University Press.
- Jacob, P. (2019). Intentionality. En *The Stanford Encyclopedia of Philosophy* (Winter 2019). <https://plato.stanford.edu/archives/win2019/entries/intentionality/>
- Locke, J. (1999). *Ensayo sobre el entendimiento humano* (E. O’Gorman, Trad.; Segunda edición en español). Fondo de Cultura Económica.
- Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, 83(4), 435. <https://doi.org/10.2307/2183914>
- Peirce, C. (2020). *Charles S. Peirce. Selected Writings on Semiotics, 1894–1912*. De Gruyter Mouton. <https://doi.org/10.1515/9783110607390>
- Piccinini, G. (2010). The Mind as Neural Software? Understanding Functionalism, Computationalism, and Computational Functionalism. *Philosophy and Phenomenological Research*, 81(2), 269-311.
- Piccinini, G. (2016a). The Computational Theory of Cognition. En V. C. Müller (Ed.), *Fundamental Issues of Artificial Intelligence* (pp. 203-221). Springer International Publishing. [https://doi.org/10.1007/978-3-319-26485-1\\_13](https://doi.org/10.1007/978-3-319-26485-1_13)

- Piccinini, G. (2016b). The Computational Theory of Cognition. En *Fundamental Issues of Artificial Intelligence* (pp. 203-221). Springer International Publishing. [https://doi.org/10.1007/978-3-319-26485-1\\_13](https://doi.org/10.1007/978-3-319-26485-1_13)
- Priestley, M. (2011). Babbage's Engines. En M. Priestley, *A Science of Operations* (pp. 17-51). Springer London. [https://doi.org/10.1007/978-1-84882-555-0\\_2](https://doi.org/10.1007/978-1-84882-555-0_2)
- Quine, W. V. (2013). *Word and object* (New ed). MIT Press.
- Russell, B. (1905). On Denoting. *Mind, Vol. 14*.
- Schanck, R., & Abelson, R. (1977). *Scripts, Plans, Goals, and Understanding* (Lawrence Erlbaum).
- Searle, J. (1980). Minds, Brains and Programs. *Behavioral and Brain Sciences*.
- Searle, J. (1983). *Intentionality, an essay in the philosophy of mind*. Cambridge University Press.
- Searle, J. (1984). *Minds, brains, and science*. Harvard University Press.
- Searle, J. (1999). *Mind, language, and society: Philosophy in the real world* (1. paperb. ed). Basic Books.
- Searle, J. R. (2002). Twenty-one years in the chinese room. En J. M. Preston & J. M. Bishop (Eds.), *Views Into the Chinese Room: New Essays on Searle and Artificial Intelligence*. Oxford University Press.
- Turing, A. M. (1950a). Computing Machinery and Intelligence. *Mind*, 59(236), 433-460. JSTOR.
- Turing, A. M. (1950b). COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, LIX(236), 433-460. <https://doi.org/10.1093/mind/LIX.236.433>
- Turing, A. M. (1951). *Can Digital Computers Think?* TS with AMS annotations of a talk broadcast on BBC Third Programme; The Turing Digital Archive. <http://www.turingarchive.org/browse.php/B/5>
- Von Eckardt, B. (1996). *What is cognitive science?* (5. print). MIT Press.