Resumen

Dentro del marco de la filosofía de la mente se han elaborado diferentes enfoques teóricos que han intentado explicar el fenómeno de la mente humana, cómo funciona, cómo se dan en nosotros los estados mentales, a qué responden estos estados y en qué lugar se encuentra la mente. Uno de los enfoques teóricos que ha despertado gran interés en los últimos años es el que se conoce como *inteligencia artificial*. De acuerdo con este enfoque, la mente humana se puede concebir como un sistema de procesamiento de la información que opera o funciona de manera muy similar a una computadora. Así, la explicación de la conciencia o del conjunto de actividades mentales que se llevan a cabo en el cerebro y que nos permiten tener experiencias conscientes, se debe realizar a través de la aplicación de un modelo computacional, puesto que la mente humana opera de la misma manera que un computador.

Sin embargo, este enfoque computacional ha sido criticado. Debido que aun cuando seamos capaces de producir artificialmente una máquina que posea un sistema nervioso, con sinapsis neuronal y demás elementos semejantes a los nuestros, ésta no podría producir intencionalidad ni estados conscientes o mentales, ya que los estados mentales no son una cuestión de cómputo o programas artificiales adecuadamente diseñados, sino que son el producto de un proceso biológico que tiene lugar en nosotros como organismos biológicos.

Palabras claves: sistema computacional, intencionalidad, consciencia, cerebro, mente, estados neurobiológicos, habitación china, sintaxis, semántica, estados subjetivos internos.

FACULTAD DE CIENCIAS HUMANAS

PROGRAMA DE FILOSOFÍA

2015

CRÍTICA A LA TEORÍA COMPUTACIONAL DE LA MENTE DESDE EL ENFOQUE NEUROBIOLÓGICO DE LA CONCIENCIA DE JOHN SEARLE

SANDRA MARCELA SIMANCAS GAMARRA

Trabajo de grado presentado como requisito para optar al título de:

Profesional en Filosofía

Asesor:

GABRIEL EDUARDO VARGAS DUQUE

UNIVERSIDAD DE CARTAGENA

FACULTAD DE CIENCIAS HUMANAS

PROGRAMA DE FILOSOFÍA

2015

CRÍTICA A LA TEORIA COMPUTACIONAL DE LA MENTE DESDE EL ENFOQUE NEUROBIOLÓGICO DE LA CONCIENCIA DE JOHN SEARLE

SANDRA MARCELA SIMANCAS GAMARRA

UNIVERSIDAD DE CARTAGENA

FACULTAD DE CIENCIAS HUMANAS

PROGRAMA DE FILOSOFÍA

2015

Nota de aceptación

_____

_____

_____

_____

_____

Jurado

Cartagena de Indias, Julio de 2015

Tabla de contenido

Introducción 1

Capítulo I. Un análisis al enfoque computacional de la mente: Los argumentos de la IA fuerte entorno a la explicación de la mente humana 8

Capítulo II. El argumento de Searle: La habitación China 28

Capítulo III. Réplicas al argumento de la habitación China: La defensa de la IA fuerte a la teoría computacional de la mente 47

Capítulo IV. Consideraciones finales ¿Teoría computacional o enfoque neurobiológico de la mente? 78

Bibliografía 88

Introducción

El proyecto metodológico de Descartes de dudar de todas las cosas, en particular de las cosas materiales, con el propósito de liberar al espíritu de toda suerte de prejuicios, es decir, apartarlo de los malos hábitos epistemológicos a los que nos conducen los sentidos y, por consiguiente, descubrir la verdad (lo claro y distinto), dio lugar en la época de la modernidad al descubrimiento de la *res cogitans*. Este descubrimiento cartesiano, se pone de manifiesto en la IV parte del *Discurso*, cuando Descartes describe el comienzo de su metodología:

> "Tiempo ha que había advertido que, en lo tocante a las costumbres, es a veces necesario seguir opiniones que sabemos muy ciertas, como si fueran indudables […]; pero deseando yo en esta ocasión de ocuparme de indagar tan sólo la verdad, pensé que debía rechazar como absolutamente falso todo aquello en que pudiera imaginar la menor duda, con el fin de ver si después […] no quedaría en (mí) algo… enteramente indudable."(Descartes, 1982, p. 61)

Hasta que:

> "Resolví fingir que todas las cosas que hasta entonces habían penetrado en mi espíritu no eran más verdaderas que las ilusiones de mis sueños. Pero advertí luego que, queriendo yo pensar, de esa suerte, que todo era falso, era necesario que yo, que lo pensaba, fuera alguna cosa; y observando que esta verdad: ‹‹yo pienso, luego soy››, era tan firme y segura que las más extravagantes suposiciones de los escépticos no son capaces de conmoverla, juzgué que podía recibirla sin escrúpulo como primer principio de la filosofía que yo buscaba."(Descartes, 1982, p. 62)

Así, tal como afirma Campos Roldán (Roldán, 2001, p. 43), la verdad del “yo pienso, luego soy” de Descartes los es por su claridad y distinción. Éstas dos, claridad y distinción reunidas, son para él, criterio de verdad: *“las cosas que concebimos muy clara y distintamente son todas verdaderas.”*(Descartes, 1982, p. 64)

De esta manera, Descartes llega a la conclusión de que no es más que una cosa que piensa, que duda, que entiende, que niega, que afirma, que quiere, que siente e imagina. En otras palabras, un espíritu. Así pues, la proposición cartesiana “yo soy, yo existo” es una verdad evidente obtenida a través de la introspección, pues *“quizás ocurriese que si yo cesara de pensar, cesaría al mismo tiempo de existir.”*(Descartes, 1977, p. 14)

Para Descartes, el que él pueda dudar de todas las cosas materiales o referidas a la naturaleza corpórea o extensa, esto es, de las cosas cuya existencia aun le son desconocidas, presupone su existencia como una cosa que piensa, como una *res cogitans.* Pues el dudar, negar, sentir, percibir, querer etc., son todos ellos propiedades del pensamiento.

En este sentido, para Descartes (1977, p. 14), el conocimiento de sí mismo no puede depender de cosas cuya existencia aún le son desconocidas y, por consiguiente, de ninguna de las que son fingidas e inventadas por la imaginación. Puesto que imaginar es para él contemplar la figura o imagen de una cosa corpórea, y todas las cosas referidas a la naturaleza del cuerpo pueden ser sueños y quimeras inventadas por la

imaginación. Por tanto, Descartes señala (1977, p. 14), que es preciso apartar el espíritu de esa manera de concebir para que pueda conocer con distinción su propia naturaleza, pues *"nada de lo que puedo comprender por medio de la imaginación pertenece al conocimiento que tengo de mí mismo*"(Descartes, 1977, p. 14)

De esta manera, la naturaleza del alma es para Descartes mucho más cognoscible que la naturaleza del cuerpo, en la medida que su existencia como res cogitans se capta inmediatamente por la naturaleza misma del pensamiento. En otras palabras, la naturaleza del alma es concebida clara y distintamente con los ojos del entendimiento como una cosa que piensa: *"nada hay que me sea más fácil de conocer que mi propio espíritu"*.(Descartes, 1977, p. 17)

Ahora bien, para Descartes la naturaleza del alma (res cogitans) no sólo es más cognoscible que la naturaleza del cuerpo (res extensa), sino que además, éstas, son sustancias radicalmente diferentes. Es decir, hay una diferencia mente-cuerpo*: "en tanto soy sólo una cosa que piensa (…), es manifiesto que yo (mi mente) soy distinto en realidad de mi cuerpo, y que puedo existir sin él"*.(Descartes, 1977, p. 45) Y advirtió que *"hay una diferencia entre el alma y el cuerpo en el hecho de que el cuerpo sea siempre divisible por naturaleza y el alma enteramente indivisible".*(Descartes, 1977, p. 50)

Pues para Descartes (1977, p. 50), aunque el espíritu parece estar unido a todo el cuerpo, sabemos que no por ello se le quita algo a éste cuando se separa del cuerpo un pie, un brazo, o alguna otra parte. Además, no pueden llamarse en términos cartesianos partes del espíritu las facultades de querer, sentir, pensar, percibir, etc., puesto que un solo y mismo espíritu es quien piensa, quiere, siente, percibe, etc. En lo que respecta a las cosas corpóreas o extensas, ocurre lo contrario, debido a que no hay ninguna que no se pueda dividir fácilmente y, por lo tanto, que pueda entenderse como indivisible *"lo cual bastaría para enseñarme que el espíritu es por completo diferente del cuerpo".*(Descartes, 1977, p. 50)

Ahora bien, si mente y cuerpo son sustancias radicalmente diferentes, entonces ¿Cómo pueden sustancias radicalmente distintas interactuar y tener efectos la una sobre otra? La respuesta que ofrece Descartes a este dualismo es que la interacción entre la mente y el cuerpo se da a partir de la agitación de la glándula pineal, al ser ésta movida por los espíritus animales. Descartes entiende por espíritus animales:

> "Cuerpos que no tienen otra propiedad que la de ser cuerpos muy pequeños que se mueven con mucha velocidad […], de tal manera, que en ningún lugar se detienen, y a medida que algunos de ellos entran en el cerebro, salen de él otros por los poros que hay en la sustancia de éste, y estos poros les conducen a los nervios, y de éstos a los músculos, y de este modo mueven el cuerpo de todas las maneras que puede ser movido."(Descartes, 2009, p. 32)

En otras palabras, hay en el cerebro una pequeña glándula en la que el alma ejerce sus funciones, debido a que ésta, es la única parte del cerebro que no se duplica. Pues las demás partes que constituyen al cerebro son

todas dobles, como también son dobles todos los órganos de nuestros sentidos externos, en cuanto tenemos dos ojos, dos oído, dos manos:

> "La razón que me persuade de que el alma no puede tener en todo el cuerpo otro lugar en qué ejercer sus funciones, que esta glándula, es que veo que las demás partes de nuestro cerebro son todas dobles, como también tenemos dos ojos, dos manos, dos oídos, siendo dobles además todos los órganos de nuestros sentido exteriores."(Descartes, 2009, p. 47).

Esta pequeña glándula está:

> "(…) de tal modo suspendida sobre el conducto por el cual se comunican los espíritus de sus cavidades anteriores del cerebro con los de la posterior, que los meros movimientos que en ella se verifican bastan para cambiar el curso de estos espíritus y alterar los movimientos de dicha glándula."(Descartes, 2009, p. 46)

Y que por su naturaleza el alma *"recibe tantas diferentes percepciones como diversos movimientos se producen en esta glándula"* y que *"puede también ser diversamente movida por el alma", que a su vez "impulsa a los espíritus que hacen mover el cuerpo"*(Descartes, 2009, p. 49).

De este modo, afirma Descartes (2009, p. 41), cuando los objetos de nuestros sentidos ocasionan algunos movimientos en los órganos de los sentidos exteriores, producen también en el cerebro, por medio de los nervios, movimientos que hacen que el alma los sienta, logrando el cuerpo material mover al alma inmaterial.

Ejemplo de esto son la percepciones que referimos a nuestro cuerpo o a algunas de sus partes, tales como la sed, el hambre, el dolor, el calor, el frio y otros apetitos naturales nuestros, producidos por la agitación de la glándula pineal al ser movida por los espíritus animales.

Mientras que el alma inmaterial logra mover al cuerpo material a la acción cuando mueve diversamente la glándula pineal, logrando impulsar *"los espíritus animales que la rodean hacia los poros del cerebro, que los conducen a los músculos a través de los nervios, medio por el cual la glándula les hace dar movimientos a los miembros"*(Descartes, 2009, p. 49).

En otras palabras, la agitación de la glándula pineal por parte de los diversos movimientos del alma, impulsa a los espíritus animales a mover el cuerpo material de todas las maneras posibles. Es así como logran interactuar, según Descartes, la mente y el cuerpo.

Esta postura cartesiana del interaccionismo dualista, dio lugar en la época de la modernidad al problema mente-cuerpo, que contemporáneamente se ha llegado a desarrollar en la inteligencia artificial. Es precisamente esta última postura la que pretendo criticar desde el enfoque neurobiológico de la conciencia propuesto por John Searle. Es decir, el propósito del presente trabajo es mostrar cuáles son las falencias que subyacen en una teoría computacional de la mente y sus procesos, como la que propone la inteligencia artificial fuerte.

Para ello dividiré mi trabajo en cuatro capítulos. En el primero me ocuparé de presentar los argumentos propuestos por la IA fuerte en torno a la explicación de la mente humana. Según este enfoque, los procesos mentales son procesos computacionalmente definidos, por lo que cualquier sistema físico que tuviese el programa correcto con los *inputs* y los *outputs*

correctos tendría una mente en exactamente en el mismo sentido que los seres humanos.

En el segundo capítulo, me ocuparé de presentar el argumento de la habitación china propuesto por Searle. Esto con el propósito de mostrar que el manejo formal de símbolos no es suficiente para la atribución de intencionalidad en las máquinas, tal como afirman los teóricos de la IA fuerte.

En el tercer capítulo, presentaré las distintas réplicas que se han elaborado en torno al argumento de la habitación china y las respuestas que nos ofrece Searle a éstas. Adicionalmente, algunas críticas que han surgido actualmente tales como: la imposibilidad del supuesto de la sustitución y el problema de las otras mentes, al que nos conduce el argumento de la habitación china.

En el cuarto capítulo, me ocuparé de resumir los dos enfoques teóricos de la mente que han sido objeto de análisis en este trabajo. El propósito aquí es mostrar cuál de estas posturas, desde mi punto de vista, es la que nos ofrece una mejor explicación de la conciencia.

Capítulo I

Un análisis al enfoque computacional de la mente: los argumentos de la IA fuerte entorno a la explicación de la mente humana.

Dentro del marco de la filosofía de la mente se han elaborado diferentes enfoques teóricos que han intentado explicar el fenómeno de la mente humana, cómo funciona, cómo se dan en nosotros los estados mentales, a qué responden estos estados y en qué lugar se encuentra la mente.

Uno de los enfoques teóricos que ha despertado gran interés en los últimos años es el que se conoce como *inteligencia artificial*. De acuerdo con este enfoque, la mente humana se puede concebir como un sistema de procesamiento de la información que opera o funciona de manera muy similar a una computadora.

Así, la explicación de la conciencia o del conjunto de actividades mentales que se llevan a cabo en el cerebro y que nos permiten tener experiencias conscientes, se debe realizar a través de la aplicación de un modelo computacional, puesto que la mente humana opera de la misma manera que un computador. Éste último al igual que la mente, es capaz ejecutar ciertas actividades cognitivas y operativas muy avanzadas. El

objetivo de la inteligencia artificial emular por medio de máquinas o dispositivos electrónicos tantas actividades mentales como sea posible, y mejorar las que llevan a cabo los seres humanos.

De acuerdo con Penrose (1996, pp. 18-19), el interés por los resultados de la inteligencia artificial procede al menos de cuatro direcciones. La primera es el estudio de la robótica, que está interesada en la aplicación de los dispositivos mecánicos que pueden realizar tareas inteligentes que anteriormente sólo podían ejecutar los seres humanos, y realizarlas con una velocidad y precisión superiores a la de cualquier ser humano.

La segunda es el desarrollo de los llamados *sistemas expertos*, con los que se intenta codificar el conocimiento esencial de toda profesión: medicina, abogacía, etc., en un paquete de ordenador. Otorgándole a la cuestión de si las computadoras pueden mostrar o imitar inteligencia auténtica importantes implicaciones sociales.

La tercera es la psicología, que confía en que tratando de imitar a través de dispositivos electrónicos el comportamiento del cerebro humano y de las distintas cualidades o actividades mentales llevadas a cabo en éste, se puedan aprender cosas importantes sobre el funcionamiento del cerebro humano.

La cuarta es la esperanza de que la IA tuviera algo que decir sobre cuestiones profundas de la filosofía y que nos proporcionara algunos elementos nuevos del concepto *mente*.

Dentro del marco de la inteligencia artificial se han realizado avances en la elaboración tecnológica de dispositivos electrónicos capaces de simular inteligencia e intencionalidad, y en cierta medida con la capacidad de adoptar comportamientos similares a los seres humanos. Penrose señala que uno de los primeros dispositivos de inteligencia artificial fue la tortuga de W. Grey Walter, construida a comienzos de los años cincuenta (1996, p. 19). Este dispositivo se movía por el suelo hasta agotar sus baterías. Posteriormente, ella misma se conectaba a la energía eléctrica con el objetivo de recargarse. Una vez se recargaba la tortuga se desconectaba por sí misma y volvía a su aventura por el suelo.

Otro ejemplo que nos brinda Penrose es el programa de computadora de Terry Winograd diseñada en 1972. Dicho artefacto tiene la capacidad de conversar sobre lo que hacía con una colección de bloques de varias formas y colores, y que por medio de simulación colocaba unos sobre otros en diferente ordene y disposición.

De acuerdo con Penrose, las computadoras que juegan ajedrez proporcionan los mejores ejemplos de máquinas que poseen lo que puede ser considerado conducta inteligente. Pues pareciese que en la realización de cada jugada asumieran posturas intencionales que denotan en ellas inteligencia y racionalidad práctica, e incluso creencias y deseos o funciones preferenciales respecto al juego. Además, han podido derrotar a algunos jugadores profesionales de ajedrez:

> “Las computadoras que juegan ajedrez proporcionan los mejores ejemplos de máquinas que poseen lo que podría ser considerado “conducta inteligente”. De hecho, algunas han alcanzado hoy día (1989) un nivel de juego más que respetable en relación con los jugadores humanos, acercándose al de “Maestro Internacional” (las puntuaciones de estas computadoras estarían por debajo de 2 300; en comparación con la del campeón mundial Kasparov, que está por encima de 2 700. Aún más impresionante es “Deep Thought” (pensamiento profundo), programado fundamentalmente por Hsiung Hsu, de la Universidad de Carnegie Mellon, y que tiene una puntuación cercana a 2500 Elo, y recientemente logró la notable proeza de compartir el primer puesto (con el Gran Maestro Tony Miles) en un torneo de ajedrez.”(Penrose, 1996, p. 20)

Penrose afirma que una de las pretensiones de la inteligencia artificial es proporcionar una vía hacia el entendimiento de las cualidades mentales, tales como la felicidad, el dolor o el hambre, entre otros (1996, p. 21). Toma como ejemplo la tortuga de Grey Walter y sostiene que cuando sus baterías están bajas su pauta de comportamiento cambia, entonces ésta actúa de la misma manera en la que actuaría un ser humano o cualquier otro animal cuando siente hambre:

> “No sería un grave abuso de lenguaje decir que la tortuga de Grey Walter está “hambrienta” cuando actúa de esta forma. Algún mecanismo interno es sensible al estado de carga de su batería, y cuando éste caía por debajo de cierto nivel, orientaba a la tortuga hacia una pauta de comportamiento diferente. Sin duda existe una operación similar en los animales cuando empiezan a tener hambre, sólo que los cambios de comportamiento son más complicados y sutiles.” (Penrose, 1996, p. 21)

De este modo, para los defensores de la inteligencia artificial los conceptos mentales de dolor, felicidad o paciencia, pueden modelarse adecuadamente de la misma forma en la que se modeló el concepto de “hambre” en el ejemplo de la tortuga Grey Walter. Pues para los defensores

de este enfoque, cualquier dispositivo electrónico puede ser capaz de simular estados mentales a través de la realización de cómputos eficientes.

Este punto de vista fue expuesto vigorosamente por Alan Turing en su famoso artículo *"La maquinaria de computación y la inteligencia".* En este artículo, Turing formula la pregunta ¿pueden las máquinas pensar? Señala que la respuesta a la pregunta no estriba en una definición a priori de qué significan las palabras "pensar" y "máquina", sino en preguntar si alguna computadora imaginable podría llegar a participar en el juego de las imitaciones (1994, p. 53).

La prueba de Turing consiste en que la computadora y algún voluntario humano se ocultan de la vista de un interrogador o examinador. Mediante el simple procedimiento de plantear preguntas de prueba a cada uno de ellos debe poder distinguir entre la computadora y el ser humano. Tanto preguntas como respuestas se trasmiten de modo impersonal. Por ejemplo, pulsadas en un teclado o mostradas en una pantalla. Al interrogador no se le da más información que la que obtiene de esa secuencia de preguntas y respuesta. El ser humano responde a las preguntas sinceramente y trata de persuadir al interrogador de que él es realmente el ser humano. Por otro lado, la computadora está programada para mentir y tratar de convencer al interrogador de que ella es el humano.

Ahora bien, si en el curso de la prueba la interrogadora es incapaz de identificar al ser humano, se considera que la computadora ha superado la prueba.

Frente a esta prueba, Turing formula la pregunta: ¿Basta el resultado del juego de las imitaciones para poder atribuirle a las computadoras conducta inteligente? A la cual el propio Turing da una respuesta afirmativa.

Turing confía en que tras los avances tecnológicos se podrían desarrollar programas formales que permitan la simulación completa de inteligencia e intencionalidad. Cree que se pueden producir en las máquinas comportamientos cognitivos idénticos los humanos y que, por consiguiente, las máquinas pueden responder al juego de las imitaciones o cualquier otro juego humano de cognición.

Con ello, Turing pretende señalar el camino que permita establecer que la mente se puede emular por completo con dispositivos electrónicos más sofisticados. De acuerdo con Turing, tales dispositivos poseen las mismas propiedades relevantes que posee la mente humana. Sólo es cuestión de tiempo para que la tecnología se desarrolle al punto de que nos sea imposible distinguir una mente humana de una mente computacional.

Para Turing se podrían elaborar sistemas completos de inferencia lógica integrados a la máquina, y que sirvan como cómputos o diseños de aprendizaje que le permitan a la computadora operar de acuerdo a órdenes, simulando con ello conducta inteligente:

> "La máquina debería construirse de tal manera que tan pronto como se clasifique una proposición imperativa como "bien establecida" ocurra automáticamente la acción adecuada. Para ilustrar esto, supongamos que el profesor le dice a la máquina: "haz tu tarea escolar ahora". Esto puede causar que el enunciado "el profesor dice: haz tu tarea escolar ahora" se incluya entre los hechos bien establecido…Las proposiciones que conducen a imperativos de este tipo podrían ser: "cuando se mencione a Sócrates, utiliza el silogismo en Bárbara" o "si un método ha demostrado ser más rápido que otro, no utilices el método más lento". Algunos de ellos pueden ser "dados por una autoridad", pero otros quizás sean producidos por la propia máquina, por inducción científica, por ejemplo". (Turing, 1994, pp. 78-79).

De esta manera, vemos que la idea de que las máquinas sólo pueden hacer lo que se les ordena resulta totalmente extraña para Turing. En la medida en que para éste la mayoría de los programas que podemos introducir en la máquina podrían ocasionar que ésta ejecute comportamientos por sí misma, sin que dichos comportamientos hayan sido programados. Es decir, se podrían elaborar programas formales que permitan a la máquina ejecutar por sí mismas formas de conductas inteligentes e intencionales, sin que tales conductas hayan sido programadas mecánicamente por un programador:

> "El punto de que "la máquina solamente puede hacer lo que sabemos cómo ordenarle que haga" resulta extraño frente a esto. La mayoría de los programas que podemos introducir en la máquina ocasionarían que haga algo que puede no tener sentido alguno para nosotros o que nos parecerá un comportamiento totalmente aleatorio." (Turing, 1994, p. 79)

Ahora bien, Penrose señala (1996, p. 23-24) que el punto de vista conocido como *inteligencia artificial fuerte*[1], adopta una posición extrema.

[1]De acuerdo con Searle, existen dos posturas dentro del enfoque computacional de la mente, la IA débil y la IA fuerte. Para la primera, las computadoras son sólo herramientas para el estudio y la explicación de la mente humana, mientras que para la segunda, las computadoras son en sí mismas explicaciones de la cognición humana, ya que poseen cognición.

Según la inteligencia artificial fuerte los dispositivos electrónicos que acabamos de ver tales como la máquina de Turing, la tortuga de Grey Walter, el programa de computadora de Terry Winograd y las computadoras que juegan ajedrez, entre otros, no sólo son inteligentes y tienen una mente, sino que al funcionamiento lógico de cualquier dispositivo computacional se le puede atribuir un cierto tipo de cualidades mentales, incluso a los dispositivos mecánicos más simples como un termostato. En otras palabras, lo que sostiene el punto de vista de la IA fuerte es que tener una mente no es nada más que poder realizar ciertas operaciones formales. Así, la mente es como el software que opera en el hardware.

Según la inteligencia artificial fuerte los procesos mentales son procesos computacionales definidos. Es decir, consisten simplemente en una secuencia bien definida de operaciones, frecuentemente llamada algoritmo.

De esta manera, todas las actividades mentales como el pensamiento, los sentimientos, la inteligencia, la comprensión, la consciencia y demás procesos subjetivos internos que denotan intencionalidad y cognición, son simplemente algoritmos que ejecuta el cerebro, y que sólo se diferencian de los algoritmos que ejecutan los dispositivos electrónicos, en cuanto poseen mayor orden de estructura y su operación es mucho más compleja:

> “En el caso del termostato el algoritmo es extremadamente simple: el dispositivo registra si la temperatura es mayor o menor que la establecida, y a continuación dispone que el circuito se desconecte o se conecte, según el caso. Para cualquier tipo

> importante de actividad mental en el cerebro humano el algoritmo tendría que ser muchísimo más complicado pero, según el punto de vista de la IA fuerte, un algoritmo complejo diferirá enormemente sólo en el grado del sencillo algoritmo del termostato, pero no habrá diferencia de principio. Así, según la IA fuerte, la diferencia entre el funcionamiento esencial del cerebro humano (incluyendo todas sus manifestaciones conscientes) y el de un termostato radica sólo en que el primero posee una mucho mayor complicación (o quizás mayor "orden de estructura" o "propiedades auto-referentes, u otro atributo que pudiéramos asignar a un algoritmo)." (Penrose, 1996, p. 24)

Para la inteligencia artificial fuerte los algoritmos que ejecuta el cerebro podrán ser emulados por cualquier tipo de computadora o dispositivo electrónico cuando se elaboren a través de los avances de las ciencias de la tecnología dispositivos con espacio de almacenamiento y velocidad de operación idénticos a los del cerebro humano. De modo que para los defensores de la inteligencia artificial fuerte la diferencia de los algoritmos entre una máquina y la mente humana es solo una diferencia temporal, que irá desapareciendo con el diseño de nuevos programas muchos más avanzados que los elaborados hasta ahora:

> "Un algoritmo que pretenda igualar el que se presume está operando en el cerebro humano tendría que ser algo prodigioso. Pero si existiera un algoritmo de esta especie para el cerebro- y los defensores de la IA fuerte afirmarán que ciertamente sí existe- entonces podría en principio funcionar en una computadora.
>
> De hecho podría funcionar en cualquier computadora electrónica moderna de tipo general si no fuera por limitaciones de almacenamiento y velocidad de operación. Se prevé que cualquiera de estas limitaciones habrá quedado superada en las grandes y rápidas computadoras de un futuro no muy lejano. Los defensores de la IA fuerte alegarán que, donde quiera que funcione, el algoritmo experimentará autónomamente sentimientos y tendrá una conciencia. Será la mente." (Penrose, 1996, p. 24)

Esta idea de los defensores de la inteligencia artificial fuerte de desarrollar dispositivos electrónicos capaces de adoptar conductas

inteligentes y, por consiguiente, poseer procesos mentales y cognitivos similares a los de los seres humanos normales, es criticada duramente por Searle en su texto *Mentes, cerebros y programas.* Para ello Searle considerará, por cuestiones de familiaridad, el programa de un teórico de la inteligencia artificial: Roger Schank.

De acuerdo con Searle, el propósito del programa de Schank consiste en simular la capacidad humana de comprender relatos y, en consecuencia, la capacidad cognitiva de responder preguntas acerca del relato, aun cuando no toda la información que proporciona el relato se establezca explícitamente:

> "Así pues, por ejemplo, suponga que escucha la siguiente historia: "Un hombre entró a un restaurante y ordenó una hamburguesa. Cuando se la sirvieron, estaba totalmente quemada, así que el hombre estalló en cólera y abandonó el restaurante furioso, sin pagar la hamburguesa ni dejar propina." Ahora bien si le preguntamos: "¿Se comió el hombre la hamburguesa?" Usted probablemente respondería: "No, no se la comió". (Searle, 1994, p. 83)

Según Searle (1994, p. 83), para los teóricos de la inteligencia artificial fuerte, la capacidad de preguntar y responder de la máquina no sólo simula una habilidad cognitiva, sino que también podría decirse, en primer lugar, que la máquina comprende el relato y, en consecuencia, es capaz de proporcionar respuestas a las preguntas que se le hagan en relación a éste. En segundo lugar, que lo que la máquina y su programa de cómputo hacen es explicar la capacidad humana de comprender el relato y responder preguntas acerca de él.

Sin embargo, Searle hace una crítica a esta idea con el argumento de la habitación china. Este argumento consiste en que imaginemos, en primer lugar, que las historias son contadas en chino y que todas las operaciones del algoritmo de la computadora para este ejercicio concreto se suministran en inglés como un conjunto de instrucciones para manipular fichas con símbolos chinos en ella. Searle se imagina así mismo haciendo todas las manipulaciones en el interior de una habitación cerrada. Las secuencias de símbolos que representan primero las historias y luego las preguntas, se introducen en la habitación a través de una ranura pequeña, ya que no se permite ninguna otra información del exterior de la habitación.

Finalmente, cuando se han completado todas las manipulaciones, la secuencia resultante se entrega a través de la ranura. Puesto que todas estas manipulaciones simplemente ejecutan el programa de Schank, el resultado final será el equivalente chino de *sí* o *no*, según sea el caso, con las que se responderá una pregunta formulada en chino acerca de la historia narrada. Searle deja en claro que él no entiende una sola palabra de chino, de modo que no tiene idea de lo que cuentan las historias.

El punto de Searle es que la ejecución correcta de algoritmos no implica, en sí mismo, que haya tenido lugar la comprensión. Pues la ejecución de un programa por un ordenador no es suficiente para aplicarle predicados mentales. Lo que hay es una manipulación sintáctica de símbolos formales, pero no semántica:

> “En el caso del chino tengo todo lo que la inteligencia artificial puede poner dentro de mi mediante un programa, y no entiendo nada; en el caso del inglés entiendo todo y hasta el momento no existe absolutamente ninguna razón para suponer que mi comprensión tiene que ver con programas de computadora, es decir, con operaciones de cómputo sobre puros elementos formalmente especificados. Mientras el programa se defina en términos de operaciones de cómputo sobre puros elementos definidos formalmente, lo que el ejemplo sugiere es que éstos por sí solos no guardan ninguna relación interesante con la comprensión.” (Searle, 1994, p. 86)

En otras palabras, la tesis del argumento de Searle, es que no hay posibilidad alguna de que el programa de Schank pueda saber el significado de los ideogramas por simple manipulación formal. El ordenador simplemente se ha comportado *como si* entendiera chino, lo que es posible simplemente por manipulación de símbolos conforme a reglas formales. Lo que el ordenador posee es una sintaxis, pero no una semántica. Así, según Searle, el rasgo que permite diferenciar las mentes de los ordenadores es que las primeras tienen una propiedad (la semántica)que los últimos son incapaces de tener por constar sólo de sintaxis, la cual es insuficiente para generar contenidos semánticos, tal como ocurre con el programa de Schank.

Sin embargo, al argumento de la habitación china expuesto por Searle le han surgido varias réplicas que el mismo autor se ocupa de analizar. La primera es la réplica de los sistemas. La idea aquí es que aunque la persona que está en la habitación manipulando los símbolos, no entiende chino, él es sólo la unidad de procesamiento central del sistema del computador. De acuerdo con esta réplica, es todo el sistema, incluyendo la habitación, las cestas llenas de símbolos y los anaqueles que contienen los programas y

quizás también otros elementos, tomado como una totalidad, lo que entiende chino.

No obstante, Searle señala (1985, p. 40) que esta réplica está sujeta a la misma objeción que ha venido haciendo antes. No hay ninguna manera de que el sistema pueda obtener a partir del manejo formal de símbolos comprensión o significación alguna. Puesto que la persona encerrada en la habitación, como unidad de procesamiento central, no tiene ninguna manera de averiguar lo que significa cualquiera de esos símbolos, entonces tampoco puede hacerlo todo el sistema.

La segunda, es la réplica del robot. De acuerdo con esta réplica la cognición no es únicamente una cuestión de manipulación de símbolos formales, pues esta réplica añade al programa de Schank un conjunto de relaciones con el mundo exterior. En otras palabras, agrega al programa de Schank dispositivos de percepción, de modo que el programa o la computadora no sólo acepte símbolos formales como entrada y produzca símbolos formales como salida, sino que también opere o maneje al robot, de tal manera que éste pueda percibir, caminar, desplazarse, martillar, beber y, en consecuencia, actuar de manera similar a un ser humano. Con ello se supone que es capaz de comprender y simular otros estados mentales.

Frente a esta réplica Searle responde (1994, p. 92) que la inclusión de las habilidades perceptuales y motoras, no convierte al programa de Schank en un objeto intencional. Puesto que la exhibición de entradas

sensoriales y salidas motoras conectadas a otras habilidades como caminar, martillar, desplazarse de un lugar a otro, beber, entre otras, no poseen un carácter cognitivo. En otras palabras, no responden a habilidades cognitivas como sucede en un ser humano, sino a un sistema de circuitos eléctricos y a su programa. De modo que el comportamiento del robot es puramente mecánico, responde a programas formales que lo controlan:

> "Cabe mencionar en este caso que el robot carece por completo de estados intencionales, pues sencillamente se activa como resultado de su sistema de circuitos eléctricos y de su programa. Por otra parte, al ejemplificar concretamente el programa yo no tengo estados intencionales pertinentes: todo lo que hago es seguir instrucciones para manipular símbolos formales."(Searle, 1994, p. 92)

La tercera réplica que nos presenta Searle, es la de la combinación. Ésta consiste en que nos imaginemos un robot que cuenta con una computadora en forma de cerebro alojada en su cavidad cerebral y que, además, está programada con todas las sinapsis de un cerebro humano, y su comportamiento es indistinguible del de una persona. Según esta réplica, debemos imaginar que todo esto es un sistema unificado y no sólo una computadora o dispositivo electrónico con patrones de entrada y salida. En tal caso tendríamos que atribuirle intencionalidad al sistema, de la misma manera que se le atribuye a cualquier ser humano normal.

Sin embargo, Searle argumenta(1994, p. 95) que aun cuando tuviéramos la inclinación de otorgarle al robot, en virtud de su comportamiento, estados mentales similares a los nuestros, tan pronto como supiéramos que su comportamiento no es el resultado de un proceso

cognitivo-mental, sino del manejo de un programa formal, abandonaríamos por completo la suposición de intencionalidad en las máquinas. Pues el único rasgo cognitivo de intencionalidad humana que vemos es el de la persona que se ocupa del manejo del programa del robot.

En otras palabras, sólo reconocemos como intencionales y cognitivos el comportamiento de la persona que maneja la máquina y no el comportamiento de la máquina misma, ya que el comportamiento de ésta es puramente mecánico y está determinado por las ordenes intencionalmente dadas por los agentes humanos:

> "La hipótesis de que el robot tiene mente resultaría entonces injustificada e innecesaria, puesto que ya no habría ninguna razón para atribuirle intencionalidad al robot o al sistema del cual forma parte (excepto, por supuesto, la intencionalidad del hombre en el manejo de los símbolos). El manejo de símbolos formales continúa, los datos de entrada y de salida se corresponden correctamente, pero el único foco real de intencionalidad es el hombre y él no sabe nada de los estados intencionales pertinentes; no ve, por ejemplo, lo que entra en los ojos del robot; no tiene la intención de mover el brazo del robot y no comprende ninguna de las observaciones hechas al robot o por el robot."(Searle, 1994, p. 95)

Por último, se encuentra la réplica de las otras mentes. Según esta réplica, sabemos que una persona tiene estados mentales a través de los distintos comportamientos que permiten inferir en él intencionalidad. En este sentido, dice Searle,(1994, p. 96) ante la "pregunta ¿Cómo sabemos que otra persona entiende chino u otra cosa?", respondemos "Por su comportamiento". Y si este es el caso, también tendríamos que atribuirle intencionalidad o estados mentales a una computadora que sea capaz de aprobar las pruebas de conducta como cualquier otra persona.

Sin embargo, Searle sostiene (1994, p. 96) que el problema de este argumento no consiste en explicar cómo sabemos que otras personas poseen estados cognitivos, sino, más bien, en presuponer qué es lo que les atribuyo a tales estados mentales. La respuesta que nos ofrece Searle a este problema es que *"la fuerza del argumento es que no podría tratarse precisamente de procesos de cómputo y de sus resultados, porque los procesos de cómputo y sus resultados pueden existir sin el estado cognitivo" (1994, p. 96)*. En otras palabras, lo que atribuimos a una persona son estados mentales y no programas formales como los de una computadora, puesto que las mentes poseen contenidos semánticos que los ordenadores son incapaces de generar por poseer simplemente sintaxis.

Para Searle, el que los seres humanos podamos comprender chino, inglés o cualquier otra lengua, no se debe a sistemas de cómputo, sino a que somos organismos biológicos con cierta estructura bilógica (el cerebro) en la que se producen causalmente los distintos estados subjetivos conscientes que hacen parte de la actividad mental.

De esta manera, vemos que para Searle aun cuando seamos capaces de producir artificialmente una máquina que posea un sistema nervioso, con sinapsis neuronal y demás elementos semejantes a los nuestros, ésta no podría producir intencionalidad ni estados conscientes o mentales, ya que los estados mentales no son una cuestión de cómputo o programas artificiales adecuadamente diseñados, sino que son el producto

de un proceso biológico que tiene lugar en nosotros como organismos biológicos.

Searle señala (1994, pp. 100, 101, 102) que existen varias razones que han llevado a los defensores de la inteligencia artificial fuerte a pensar que un dispositivo electrónico o una computadora diseñada adecuadamente puede producir y explicar los fenómenos mentales. La primera es la confusión acerca de la noción “procesamiento de información”.

Dentro del ámbito de las ciencias cognitivas, se considera que el cerebro humano, con su mente, realiza algo que recibe el nombre de “procesamiento de información”, y análogamente la computadora, con su programa, efectúa también “procesamientos de información”, lo que permite comparar a ésta y a su programa formal con la mente humana.

Sin embargo, Searle señala que aunque la computadora pueda simular la realización de cualquier proceso de información, este proceso no se da en el mismo sentido en el que las personas procesan información. No hay en la computadora un proceso reflexivo o autoconsciente acerca de la información que recibe, tal y como ocurre en las personas cuando reciben información:

> “La computadora no realiza “procesamiento de información” en el sentido en que las personas “procesan información” cuando reflexionan, por ejemplo, en problemas de aritmética o cuando leen y responden preguntas en torno a un relato. Más bien, lo que hace la computadora es manipular símbolos formales.”(Searle, 1994, pp. 100-101)

La segunda razón, tal como señala Searle(1994, p. 101), es que en gran parte de la inteligencia artificial hay un conductismo u operacionalismo residual. Puesto que las computadoras adecuadamente programadas pueden exhibir patrones de entrada y salida similares a los de los seres humanos, podríamos caer en la tentación de afirmar que los estados mentales de las computadoras son similares a los estados mentales humanos.

No obstante, Searle afirma que dichos patrones sensoriales y motores de entrada y salida no poseen, como sucede en las personas, intencionalidad. En otras palabras no poseen un carácter cognitivo-mental, sino mecánico, pues dichos patrones responden simplemente a programas formales:

> "Mi sumadora de escritorio tiene capacidad de calcular, pero carece de intencionalidad, y en este artículo he intentado demostrar que un sistema podría tener capacidades de entrada y salida de información que duplicaran las de un hablante nativo del chino y seguir sin entender chino, sin importar cómo estuviera programada."(Searle, 1994, p. 101)

La tercera razón que nos ofrece Searle, es que este operacionalismo o conductismo residual da lugar a una forma residual de dualismo. Según Searle (1994, p. 102), dentro del marco de la IA fuerte lo que importa son los programas, y estos son independientes de su realización en máquinas, puesto que para los defensores de la IA fuerte dichos programas podrían ser realizados por una máquina electrónica, una sustancia mental cartesiana o un espíritu del mundo hegeliano. En otras palabras, no hay una relación causal entre la computadora y los programas que ésta ejecuta, debido a que

estos programas se pueden llevar a cabo con independencia de la computadora. En este sentido, Searle afirma que la IA fuerte:

> "Está comprometida con la hipótesis de que la relación del cerebro con la conciencia no es para nada una relación causal, sino que la conciencia consiste simplemente en programas en el cerebro. Y niega que la específica neurobiología del cerebro cuente para la conciencia en particular y para la mente en general. De acuerdo con la IA fuerte, la mente y la conciencia no son procesos concretos, físicos, biológicos, como el crecer, vivir y el dirigir, sino algo formal y abstracto. En realidad, así es exactamente como caracterizaban a la mente en un libro anterior Daniel Dennett y su coautor, Douglas Hofstadter. Se trata, decían, ‹‹de un tipo abstracto de cosa, cuya identidad es independiente de cualquier corporeización física particular››. En esta idea se expresa el dualismo residual típico de la teoría computacional de la mente."(Searle, 1997, pp. 171-172)

Al contrario del argumento de los defensores de la inteligencia artificial fuerte, para Searle sí hay una relación causal entre cerebro y mente. Puesto que la mente o la conciencia emerge biológicamente del sistema o la estructura cerebral. Es decir, necesitamos de un cerebro para que se puedan dar en nosotros estados subjetivos conscientes o neurológicos. Debido a que es precisamente en la estructura del cerebro dónde se llevan a cabo los procesos neurobiológicos que causan nuestros estados subjetivos conscientes.

De esta manera, vemos que los defensores de la inteligencia artificial fuerte, fracasan, de acuerdo con Searle, en su intento de explicar los fenómenos mentales a través de procesos computacionales, ya que la actividad mental es un rasgo biológico del ser humano, y no el producto del manejo de símbolos formales o de programas elaborados para el manejo de tales símbolos. En otras palabras, los estados mentales son fenómenos

biológicos, pues la conciencia la intencionalidad y la causación mental son todas ellas parte de la historia de nuestra vida biológica, junto con el crecimiento, la reproducción, la secreción de bilis y la digestión, entre otros fenómenos biológicos que se dan en nosotros.

En el siguiente capítulo me ocuparé de formular y explicar detalladamente el argumento de la habitación china, con el propósito de mostrar cuáles son las falencias que subyacen en una teoría computacional de la mente y sus procesos, como la que propone la inteligencia artificial fuerte.

Capítulo II

El argumento de Searle: La habitación china.

He presentado en el capítulo anterior los argumentos de la IA fuerte en lo que respecta a la explicación de la mente y sus procesos. Me ocuparé ahora de explicar detalladamente el argumento de la habitación china con el propósito de mostrar que el manejo formal de símbolos no es suficiente para la atribución de intencionalidad en las máquinas. Para ello, comenzaré por presentar un pequeño resumen de la postura de la IA fuerte.

Posteriormente formularé la versión modificada de la habitación china de Searle, con el fin de resaltar que la sintaxis no es un rasgo intrínseco de los sistemas computacionales como parecen presuponer los defensores de la IA fuerte, sino que se derivan de la interpretación que hacen los agentes conscientes.

De acuerdo con la IA fuerte el cerebro es solamente un computador digital y la mente es solamente un programa de computador. Es decir, la mente es al cerebro lo que el programa es al hardware del computador. Para la IA fuerte no hay nada esencialmente biológico por lo que respecta a la mente humana, pues el cerebro es uno de un número indefinidamente extenso de diferentes géneros de hardware de computador que podrían

servir de soporte a los programas formales que constituyen la inteligencia humana.

Según la IA fuerte cualquier sistema físico que tuviese el programa correcto con los *inputs* y los *outputs* correctos tendría una mente en exactamente el mismo sentido que los seres humanos. En otras palabras, al funcionamiento normal de cualquier dispositivo computacional se le pueden atribuir cualidades mentales. Incluso, hasta a los dispositivos mecánicos más simples como, por ejemplo un termostato. Desde esta perspectiva, cualquier sistema que sea capaz de manipular símbolos de una manera correcta exhibe inteligencia en el mismo sentido que los seres humanos conscientes la exhiben:

> "Según este punto de vista, cualquier sistema físico que tuviese el programa correcto con los inputs y los outputs correctos tendría una mente en exactamente en el mismo sentido que tú y yo tenemos mente. Así, por ejemplo, si se hiciese un computador con viejas latas de cerveza y se le suministrase energía por medio de molinillos de viento, si tuviera el programa correcto, tendría que tener una mente. Y el punto no es que, dado todo lo que sabemos, podría tener pensamientos y sensaciones, sino más bien que tiene que tener pensamientos y sensaciones, puesto que todo aquello en lo que consiste tener pensamientos y sensaciones es esto: llevar a cabo el programa correcto."(Searle, 1985, p. 34)

Este punto de vista es criticado por Searle, puesto que los programas y modelos de las computadoras que utilizan los defensores de la IA fuerte en sus argumentos son puramente sintácticos. El punto de Searle es que ningún sistema es capaz, por sí mismo, de tener consciencia simplemente en virtud de la ejecución correcta de cómputos.

Así, por ejemplo, Searle señala (1997, p. 187) que hemos diseñado calculadoras de tal modo que producen símbolos que nosotros podemos interpretar como respuestas correctas a preguntas aritméticas que hemos introducido, pero que las calculadoras como tal no saben nada de números, ni de sumas.

Análogamente, ocurre lo mismo con los dispositivos computacionales que juegan al ajedrez: no saben nada de ajedrez, ni de jugadas. Para Searle, estos dispositivos que emplean los defensores de la IA fuerte, sólo son máquinas para manipular símbolos carentes de significado. Sólo somos nosotros quienes les damos interpretación y semántica a tales símbolos:

> "Nosotros somos quienes podemos interpretar los símbolos que introducimos como posiciones de ajedrez, y los símbolos que la máquina produce, como jugadas de ajedrez, porque hemos diseñado a la máquina para que lo haga así: saca símbolos sobre jugadas en respuesta a símbolos sobre posiciones. Igualmente podríamos interpretar las entradas como posiciones en un ballet, y las salidas, como cambios coreográficos. A la máquina le da lo mismo. La idea de que, de uno u otro modo, éste u otros programas son la clave de la conciencia es pura fantasía."(Searle, 1997, p. 187)

Según Searle, no podemos reducir la mente y los procesos mentales al manejo formal de símbolos abstractos, debido a que tener una mente es algo más que tener procesos formales o sintácticos. Las mentes tienen contenidos mentales y contenidos semánticos, debido a que mis pensamientos, mis creencias, y mis deseos son sobre algo, se refieren a algo o conciernen a estados de cosas del mundo. Hacen esto porque sus contenidos se dirigen a los estados de cosas del mundo:

> “Nuestros estados mentales internos tiene, por definición, ciertos tipos de contenido. Si estoy pensando en Kansas City, o deseando tener una cerveza fría para beber, o preguntándome si habrá una caída en los tipos de interés, en cada caso mi estado mental tiene un cierto contenido mental además de cualesquiera otros rasgos formales que pueda tener. Esto es, incluso si mis pensamientos se me presentan en cadenas de símbolos tiene que haber más que las cadenas abstractas, puesto que las cadenas por sí mismas no pueden tener significado alguno. Si mis pensamientos han de ser sobre algo, entonces las cadenas tienen que tener un significado que hace que sean los pensamientos sobre esas cosas. En una palabra, la mente tiene más que una sintaxis, tiene una semántica.”(Searle, 1994, p. 37)

De esta manera, para Searle, la razón por la que un programa de computador no puede jamás ser una mente, es simplemente porque un programa de computador es solamente sintáctico, y las mentes son más que sintácticas, en el sentido de que tienen algo más que una estructura formal: tienen contenido.

Para defender esta posición, Searle diseña el argumento de*la habitación china*. Con este experimento mental, Searle pretende elaborar una analogía con los mismos elementos que pone en juego la IA fuerte: programas computacionales y test de Turing. Si se puede demostrar que la ejecución de un programa informático o computacional no puede emular la semántica, entonces la tesis de la IA fuerte deja por fuera un aspecto importante para la explicación adecuada de la inteligencia humana, a saber: la consciencia. Esto es así porque parte de lo que significa pensar, consiste en ser capaz de dirigirse al mundo a través de los contenidos propios generados por la mente.

Searle nos propone que imaginemos que se encierra a una persona en una habitación y que en esa habitación hay diversas cestas llenas de símbolos chinos. Imaginemos además que la persona de la habitación no entiende chino, pero que se le da un libro de reglas en castellano para manipular esos símbolos de manera puramente formal. Así la regla podría decir: “toma un signo changyuan-changyuan de la cesta número uno y colócalo al lado de un signo chongyoun-chongyoun de la cesta número dos”.

Siguiendo con el argumento, supongamos que la persona no sabe que los símbolos introducidos en la habitación son denominados “preguntas” de las personas que se encuentran fuera de la habitación, y que los símbolos que la persona de la habitación devuelve fuera de la habitación son denominados “respuestas a las preguntas”.

En este punto, Searle señala que los programadores son tan buenos al diseñar los programas y que la persona de la habitación es tan buena manipulando símbolos que enseguida sus respuestas son indistinguibles de las de un hablante nativo del chino. La persona se encuentra encerrada en su habitación barajando sus símbolos chinos y devolviendo símbolos chinos en respuesta a los símbolos chinos que entran. Searle advierte que sobre la situación, tal y como se ha descrito, no hay manera de que la persona pueda aprender nada de chino sólo manipulando esos símbolos formales.

Es decir, que a partir de la mera organización de símbolos no es posible que alguien aprenda un idioma. Es necesario que otras funciones se

hagan presentes y en el caso de los programas informáticos tales funciones están ausentes. Esa es la razón por la cual Searle piensa que se necesitan las funciones semánticas que surgen del cerebro y posibilitan la perspectiva de primera persona con la que opera la mente consciente.

En conclusión, puesto que la persona no aprende el chino llevando a cabo un programa informático, es imposible que cualquier otro sistema computacional comprenda el chino tomando como base el mismo programa informático de manipulación formal de símbolos. La persona que ejecuta el programa no tiene nada que no posea un ordenador cualquiera, pues ambos cumplen el requisito de poder ejecutarlo sin que importe el medio físico en que se efectúa la ejecución de los símbolos formales en chino.

A partir del ejemplo de la habitación china, Searle extrae la siguiente conclusión general:

> "Lo dicho para el chino vale igual para otras formas de cognición. La mera manipulación de símbolos, no basta, por sí misma, para garantizar cognición, percepción, comprensión, pensamiento, y así sucesivamente. Y dado que los ordenadores, en su cualidad de tales, son dispositivos de manipulación de símbolos, la mera ejecución del programa no basta para garantizar la cognición."(Searle, 1990, p. 10)

Lo que Searle pretende mostrarnos, es que la ejecución correcta de algoritmos o del manejo de programas para la manipulación de símbolos no implica en sí mismo que haya tenido lugar comprensión alguna. La persona encerrada en la habitación manipula, más no comprende los símbolos. Lo que hay es un manejo formal de símbolos que no incluye contenido semántico o significado agregado a los símbolos en chino.

De este modo, todo lo que el computador tiene, al igual que la persona de la habitación, es un programa formal para manipular símbolos en chino. El computador tiene sintaxis, pero no semántica.

De acuerdo con Searle (1985, p. 39) podemos ver la fuerza de este argumento, si contrastamos aquello a lo que se parece el ser interrogado en un lenguaje en el que no tenemos conocimiento de los significados de las palabras. Searle nos propone que nos imaginemos que en la habitación china se le formulan a la persona preguntas en castellano sobre cosas tales como su edad o episodios de su vida y que la persona responde a las preguntas.

Según Searle, la diferencia entre el caso del chino y el caso del castellano, es que la persona no entiende nada de chino y, en cambio sí, entiende el castellano.

Así pues, la persona entiende las preguntas en castellano porque están expresadas en símbolos cuyos significados le son conocidos. Similarmente, cuando la persona da las respuestas en castellano, está produciendo símbolos que son significativos para él. Pero en el caso del chino la persona manipula simplemente símbolos formales y no les añade significado alguno a ninguno de los elementos del chino, como si ocurre con el castellano.

En este sentido vemos que para Searle (Searle, 1985, p. 38) comprender un lenguaje o tener estados mentales incluye tener una

interpretación o un significado agregado a los símbolos que se están manejando, y un computador no puede tener más que símbolos formales, puesto que la operación del computador se define en términos de su capacidad para organizar los símbolos de forma pre-establecida. El programa ejecuta órdenes que, de alguna manera, están contenidas en las instrucciones originales.

Tales programas no poseen, de acuerdo con Searle, contenido. Pues la sintaxis sola no implica la semántica y los computadores en tanto dispositivos computacionales poseen, por definición, solamente sintaxis, no semántica:

> "La pregunta que queríamos plantear es ésta: ¿Puede un computador digital, tal como se ha definido, pensar? Es decir: ¿Es suficiente para, o constitutivo de, pensar el instanciar o llevar a cabo el programa correcto con los inputs y los outputs correctos? Y a esta pregunta, a diferencia de sus predecesoras la respuesta es claramente "no". Y es "no" por la razón que hemos puesto de manifiesto reiteradamente, a saber: el programa del computador está definido de manera puramente sintáctica. Pero pensar es algo más que manipular signos carentes de significado, incluye contenidos semánticos significativos. A esos contenidos semánticos es a lo que nos referimos mediante significado."(Searle, 1985, p. 42)

Para Searle, si se trata realmente de un computador, sus operaciones tienen que definirse sintácticamente; mientras que la conciencia, los sentimientos, los pensamientos, las sensaciones, las emociones y todos los demás rasgos esenciales de la mente incluyen algo más que sintaxis. Por definición el computador es incapaz de duplicar esos rasgos mentales por muy poderosa que pueda ser su capacidad para simular. En la medida que

tener una mente implica tener contenidos semánticos dirigidos intencionalmente hacia un estado de cosas del mundo:

> "¿Por qué ha pensado alguien alguna vez que los computadores podrían pensar o tener sensaciones y emociones y todo lo demás? Después de todo, podemos hacer simulaciones computacionales de cualquier proceso del que pueda darse una descripción formal. Así, podemos hacer una simulación computacional del flujo de dinero en la economía española, o del modelo de distribución de poder en el partido socialista. Podemos hacer una simulación computacional de las tormentas en los términos municipales del país, o de los incendios en los almacenes del este de Madrid. Ahora bien, en cada uno de esos casos, nadie supone que la simulación computacional es efectivamente la cosa real; nadie supone que una simulación computacional de una tormenta nos deje a todos mojados, o que sea probable que una simulación computacional de un incendio vaya a quemar la casa. ¿Por qué diablos va a suponer alguien que esté en sus cabales que una simulación computacional de procesos mentales tiene efectivamente procesos mentales? Realmente desconozco la respuesta a esto, puesto que la idea me parece desde el principio, para decirlo con franqueza, completamente disparatada."(Searle, 1985, pp. 43-44)

La semántica para Searle es muy importante, puesto que permite comprender el hecho de que la mente pueda tener una relación con la realidad, de que los estados mentales sean intencionales, es decir: tengan la capacidad de ser sobre algo distinto a ellos mismos. De este modo, si sumamos la semántica a la sintaxis, podemos comprender mucho mejor la relación entre lenguaje y realidad. En otras palabras, entender de qué manera nuestras habilidades mentales nos permiten referirnos significativamente a objetos que no tienen significado.

De esta manera, Searle establece una diferencia entre tener una semántica intrínseca y tener una semántica derivada. Cualquier sistema simbólico sólo puede exhibir semántica en virtud de los estados mentales.

Es en este sentido profundo que lo sintáctico no puede hacer referencia a nada. Una mera manipulación de símbolos no puede generar estados mentales, pues la semántica es una condición necesaria para que pueda surgir el significado.

Una vez presentada la distinción entre sintaxis y semántica que se desprende de la tesis de la habitación china, se puede afirmar que dicha tesis ya generalizada es presentada de manera más esquemática por Searle en los siguientes términos:

1. Los procesos mentales que nosotros consideramos que constituyen una mente son causados por procesos que tienen lugar dentro del cerebro. Los cerebros causan las mentes.
2. La sintaxis no es suficiente para la semántica.
3. Los programas de computador están definidos enteramente por su estructura formal o sintáctica.
4. Las mentes tienen contenidos mentales, específicamente, tienen semántica.

Estas cuatro premisas nos llevan, de acuerdo con Searle, a la siguiente conclusión: que los programas o sistemas computacionales no son mentes y no son suficientes para que emerja la mente. Podemos decir que esta conclusión se sustenta en dos razones:

La primera, es que cualquier dispositivo computacional que pudiéramos construir que tuviese estados mentales equivalentes a los

estados mentales humanos, tendría que tener poderes causales equivalentes al menos a los del cerebro, y como hemos señalado reiteradamente, un computador es incapaz de producir por sí mismo una mente, ya que por definición sólo posee sintaxis y no semántica. La sintaxis por sí misma no garantiza la presencia de conciencia, debido a que la manipulación de símbolos abstractos no tiene capacidades causales para causar conciencia, porque no tiene capacidad causal alguna:

> "Lo que yo hago es más bien ofrecer una prueba de que las operaciones computacionales, por sí mismas, es decir, no son suficientes para garantizar la presencia de conciencia. La prueba ofrecida era que las manipulaciones de símbolos se definen en términos sintácticos abstractos, y que la sintaxis, por sí misma, no tiene contenido mental, consciente o de otro tipo. Además, los símbolos abstractos no tienen capacidades causales para causar conciencia, porque no tienen capacidad causal alguna. Todas las capacidades causales están en el medio de ejecución. Un particular medio en el que se ejecuta un programa, mi cerebro, por ejemplo, podría tener capacidades causales independientes para causar conciencia. Pero la operación del programa tiene que definirse de un modo totalmente independiente de cualquier medio ejecutante, pues la definición del programa es puramente formal, lo que permite su ejecución en cualesquiera medios."(Searle, 1997, p. 188)

Pues un principio básico en la filosofía de la mente de Searle, es que todos los fenómenos mentales, ya sean conscientes o inconscientes, visuales o auditivos, dolorosos, cosquillosos, toda nuestra vida mental, está efectivamente causada por procesos que acaecen en el cerebro. El cerebro causa las mentes y duplicar una mente significa haber duplicado anteriormente los poderes causales del cerebro.

La segunda, es que para Searle (1997, p. 177), el hecho de que no dispongamos de una perspectiva teórica unificada de cómo funciona el

cerebro, es decir, como se lleva a cabo el correlato causal entre nuestros estados mentales y los procesos neurobiológicos que los causan, imposibilita la idea de elaborar sistemas computacionales capaces de emular la mente:

> "Ya hacemos corazones artificiales en las fábricas. ¿Por qué habría de ser diferente en el caso de los cerebros? No hay obstáculos lógicos para el cerebro que sean mayores que los obstáculos para un corazón artificial. Ni que decir tiene que las dificultades son enormes, pero son de naturaleza práctica y científica, no lógica y filosófica. Puesto que desconocemos cómo lo hacen los cerebros reales, estamos en una pésima situación para fabricar cerebros artificiales que causen conciencia."(Searle, 1997, p. 181)

Ahora bien, hemos visto que el argumento de la habitación china de Searle nos plantea la idea de que no podemos reducir la explicación de la mente y los procesos neurobiológicos que causan biológicamente la mente, a la simple actividad de manipular símbolos formales. Pues la mente posee un contenido mental que la diferencia en tanto producto biológico de cualquier sistema que simule computacionalmente estados mentales, puesto que dicha simulación sólo se define en función del manejo de símbolos y programas formales que poseen un carácter sintáctico, no semántico.

Esta idea de que los sistemas computacionales se definen sintácticamente en términos de la manipulación de símbolos, es modificada posteriormente por Searle en su texto *"El redescubrimiento de la mente"*. En este texto, Searle plantea que la sintaxis y los símbolos no se definen en términos de la física. Aunque las instancias de un símbolo son siempre instancias físicas, símbolo y sintaxis no se definen en términos de rasgos

físicos. La sintaxis no es intrínseca a la física como piensan los defensores de la IA fuerte:

> "‹‹La sintaxis›› no es el nombre de un rasgo físico, como masa o gravedad. Por el contrario, hablan de ‹‹motores sintácticos›› e incluso de ‹‹motores semánticos›› como si tal manera de hablar fuese semejante aquella en la que se habla de motores de gasolina o motores diesel, como si pudiera ser un simple hecho objetivo que el cerebro, o cualquier otra cosa, sea un motor sintáctico."(Searle, 1996, p. 214)

Según Searle (1996, p. 229), esto tiene como consecuencia el que los sistemas computacionales no se descubren en la física, sino que se asignan a ella. Pues ciertos fenómenos físicos se usan, programan e interpretan sintácticamente.

De aquí que sintaxis y símbolos son relativos al observador. Esto se debe a que nociones tales como computación, algoritmo y programa, no nombran rasgos físicos intrínsecos de los sistemas. Un estado físico de un sistema es un estado computacional sólo de manera relativa a la asignación a ese estado de algún rol, función o interpretación computacional:

> "La sintaxis no es intrínseca a la física. La adscripción de propiedades sintácticas es siempre relativa a un agente u observador que trata como sintácticos ciertos fenómenos físicos." (Searle, 1996, p. 213)De este modo, para Searle, no hay ninguna manera en que pueda descubrirse que algo es intrínsecamente un ordenador digital, en la medida que su caracterización como ordenador digital es simplemente relativo a un observador que asigna una interpretación sintáctica a los rasgos puramente físicos del sistema.

Así, Searle señala (1996, p. 215) que la caracterización de un proceso como computacional es una caracterización de un sistema físico desde fuera, y la identificación del proceso como computacional no identifica

un rasgo intrínseco de la física, es esencialmente una caracterización completamente relativa al observador, al agente que interpreta ciertas marcas físicas o sistemas físicos como significativos:

> "Este punto ha de entenderse de manera precisa. No estoy diciendo que haya límites a priori respecto de los patrones que podemos descubrir en la naturaleza. Sin duda, podríamos descubrir un patrón de eventos en mi cerebro que fuera isomórfico con la implementación del programa de edición de textos de mi ordenador. Pero decir que algo está funcionando como un proceso computacional es decir algo más que está ocurriendo un patrón de eventos físicos. Requiera la asignación de una interpretación computacional por parte de un agente. Análogamente, podríamos descubrir objetos en la naturaleza que tuviesen la misma clase de contorno que las sillas y que, por lo tanto, pudiesen ser usadas como sillas; pero no podríamos descubrir en la naturaleza objetos que estuviesen funcionando como sillas, excepto de manera relativa a algunos agentes que las consideran o usan como sillas."(Searle, 1996, pp. 215-216)

De acuerdo con Searle (1996, p. 216), para entender este argumento es esencial entender la distinción entre rasgos del mundo que son intrínsecos y rasgos que son relativos al observador. Las expresiones "masa", "atracción gravitatoria" y "moléculas", nombran rasgos del mundo que son intrínsecos. Pues si todos los observadores y usuarios dejasen de existir el mundo contendría aún masa, atracción gravitatoria y moléculas. Pero expresiones como "bañera", "silla", y otras expresiones semejantes a éstas, no nombran, según Searle, rasgos intrínsecos del mundo, sino objetos especificando algún rasgo que se les ha asignado, rasgos que son relativos a observadores y usuarios:

> "Si no hubiese habido jamás usuario u observador alguno habría con todo montañas, moléculas, masa y atracción gravitatoria. Pero si no hubiese habido nunca ningún usuario u observador, no habría rasgos tales como ser un día para merendar al campo, o ser una silla o una bañera. La asignación de rasgos relativos al

> observador a rasgos intrínsecos no es arbitraria. Algunos rasgos intrínsecos del mundo facilitan su uso como, por ejemplo, sillas o bañeras. Pero el rasgo de ser una silla o una bañera o un día precioso para ir al campo a merendar es un rasgo que sólo existe de manera relativa a usuarios u observadores."(Searle, 1996, p. 216)

Con este argumento, Searle pretende mostrarnos, que de acuerdo con la definición estándar de computación, los programas o estructuras sintácticas, son relativos al observador, no son intrínsecos. En otras palabras, la sintaxis no es parte de la física, sino que es resultado de la interpretación por parte de un agente observador que usa y programa ciertos sistemas físicos como sintácticos.

Para Searle, el hecho de que la sintaxis no sea intrínseca a la física trae como consecuencia el que si la computación se define sintácticamente, entonces nada es intrínsecamente un ordenador digital solamente en virtud de sus propiedades físicas. De esto se sigue, que no se puede descubrir que el cerebro o cualquier otra cosa es un ordenador digital, aunque se le pueda asignar una interpretación computacional como puede hacerse con cualquier otra cosa.

En otras palabras, si la computación se define en términos de la asignación de sintaxis, entonces todo puede ser un ordenador digital, en la medida que a cualquier objeto se le podrían hacer adscripciones sintácticas. Se podría describir cualquier cosa en términos de ceros y unos.

En este sentido, Searle señala (1996, p. 230) que la afirmación de la IA fuerte: "el cerebro es un ordenador digital" no tiene un sentido claro. Pues

la pregunta ¿Es el cerebro un ordenador digital? Está mal formulada. Ahora bien, si la pregunta es ¿Podemos asignar una interpretación computacional al cerebro? La respuesta es trivialmente sí. Puesto que podemos asignar una interpretación computacional a cualquier cosa, de la misma manera como un agente observador podría asignarle una interpretación sintáctica a ciertos fenómenos en virtud de sus propiedades físicas:

> "La realizabilidad múltiple de los procesos computacionalmente equivalentes en diferentes medios físicos no es sólo una señal de que los procesos son abstractos, sino de que no son intrínseco al sistema. Depende de una interpretación desde fuera. Buscamos algunos hechos objetivos que harían computacionales los procesos cerebrales; pero dado el modo en que hemos definido la computación jamás puede haber tales hechos objetivos. No podemos por un lado, decir que cualquier cosa es un ordenador digital si podemos asignarle una sintaxis y suponer, por otro, que hay una cuestión fáctica intrínseca a su operación física que decide si un sistema natural tal como el cerebro es un ordenador digital."(Searle, 1996, pp. 214-215)

Ahora bien, si lo que se pregunta es ¿Son los procesos cerebrales intrínsecamente computacionales? La respuesta es trivialmente "no", excepto en el caso de los agentes conscientes que intencionalmente llevan a cabo computaciones. Pues, como se afirmó en el capítulo anterior, en la respuesta a la réplica de la combinación, en un sistema computacional no hay un contenido intencional intrínseco al sistema que esté funcionando intencionalmente para producir una determinada conducta.

El sistema no ejecuta por sí mismo comportamiento consciente que denote en éste la condición de cognición humana. Sólo sigue de manera mecánica reglas que responden a un programa formal que está siendo llevado a cabo por un agente humano que está siguiendo conscientemente

reglas y que, en consecuencia, tiene un contenido mental que explica intencionalmente su conducta.

Además, aunque se llegara a simular computacionalmente los estados mentales, dicha simulación no sólo carecería por completo de semántica, sino que tampoco tendría poderes causales equivalentes a los del cerebro humano pues aún, cuando su comportamiento fuese semejante al de los seres humanos, éste dependería de un agente humano que llevara a cabo la computación. En la medida en que sin un agente humano, tanto el sistema computacional como el cerebro tendrían únicamente modelos; y los modelos, como afirma Searle, no tienen poderes causales adicionales a los que tienen los agentes que los diseñan.

Adicionalmente, para Searle, la mente y sus procesos subjetivos conscientes sólo son posibles en nosotros y no en los computadores, en tanto somos seres biológicos con un cerebro en el que son causados nuestros estados conscientes internos. El cerebro es un órgano biológico específico y sus procesos neurobiológicos causan formas específicas de intencionalidad. En el cerebro hay intrínsecamente procesos neurobiológicos, y éstos causan algunas veces la conciencia.

Todas las demás atribuciones mentales, son según Searle *"o bien disposicionales, como cuando adscribimos estados inconscientes al agente humano, o son relativos al observador, como cuando asignamos una interpretación computacional a sus procesos cerebrales"*(1996, p. 230). En

otras palabras, la mente y sus contenidos mentales son rasgos esencialmente intrínsecos de los seres humanos y no de los sistemas computacionales, pues lo que estos últimos poseen son simplemente estructuras formales que dependen para su interpretación de un agente observador.

En este sentido, los sistemas computacionales no tienen poderes causales en el mismo sentido en que los tienen los procesos neurobiológicos que causan nuestros estados subjetivos conscientes, pues el programa implementado en el ordenador o dispositivo computacional, no tiene una semántica intrínseca, ya que ésta depende tal y como hemos visto, de un agente observador:

> "La tesis es que hay toda una gran cantidad de símbolos que se están manipulando en el cerebro, ceros y unos que centellean de un lado a otro del cerebro a la velocidad de la luz y que son invisibles no sólo a simple vista, sino con el más potente microscopio electrónico, y que causan la cognición. Pero la dificultad es que los ceros y los unos como tales no tienen poderes causales puesto que ni siquiera existen excepto en los ojos del observador. El programa implementado no tiene poderes causales distintos del medio que lo implementa puesto que el programa no tiene existencia real, no tiene ontología más allá del medio que lo implementa. Físicamente hablando, no hay nada que sea un ‹‹nivel de programa›› separado."(Searle, 1996, p. 220)

En el capítulo siguiente me ocuparé de analizar las distintas críticas que se le han hecho al argumento de la habitación china por parte de los defensores de la IA fuerte, para luego presentar las distintas respuestas que nos ofrece Searle a cada una de ellas. Esto con el propósito de mostrar los defectos que presenta cualquier teoría que intente explicar computacionalmente los procesos mentales y los procesos neurobiológicos

que causan y sustentan la mente, como si éstos fueran un simple proceso formal de manipulación de símbolos abstractos.

Capítulo III

Réplicas al argumento de la habitación china: la defensa de la IA fuerte a la teoría computacional de la mente.

El presente capítulo tiene como propósito presentar algunas de las distintas réplicas que se han elaborado en torno al argumento de la habitación china y la respuesta que nos ofrece Searle a éstas. Adicionalmente, presentaré algunas críticas que han surgido actualmente tales como: la imposibilidad del supuesto de la sustitución y, el problema de las otras mentes al que nos conduce el argumento de la habitación china.

Pero para ello, me ocuparé primero de presentar el concepto de intencionalidad en Searle, puesto que dicho concepto constituye la base sobre la cual labora su argumento de la habitación china.

De acuerdo con Salcedo Albarán (2004, p. 43), la intencionalidad en Searle, no consiste en algo misterioso, puesto que es tan sólo el resultado de algunos procesos químicos y biológicos propios de los seres humanos y algunos animales. Es decir, para Searle la intencionalidad es un proceso biológico idéntico a la digestión o al flujo sanguíneo. Así, la intencionalidad se interpreta como una macro-propiedad, que es el resultado del comportamiento de micro-propiedad cerebral y neuronal:

> “Los fenómenos mentales son causados por los procesos que suceden en el cerebro en el nivel neuronal o modular, pero están realizados en el mismo sistema que consiste en neuronas organizadas en módulos.” (Searle, 1995, p. 433)

En otras palabras, Searle ha propuesto considerar los fenómenos mentales como causados por y realizados en el sistema neuronal del cerebro, aplicando aquí la distinción entre macro-propiedades y micro-propiedades. De esta manera, así como la solidez de un objeto es una macro-propiedad del objeto, producida por una determinada estructura molecular que es una micro-propiedad del mismo, igualmente hay que considerar los estados mentales como macro-propiedades del cerebro producidas por ciertas micro-propiedades del mismo como es su estructura neuronal. Así pues, Searle distingue dos niveles de descripción causal en el cerebro, un macro-nivel de procesos neurofisiológicos mentales (fenómenos mentales) y un micro-nivel de procesos fisiológicos neuronales (estructura neuronal):

> “La característica de superficie F es causada por el comportamiento de los micro elementos M, y que al mismo tiempo está realizada en el sistema de los micro elementos. Las relaciones entre M y F son causales pero al mismo tiempo F es, simplemente, una característica de nivel más elevado del sistema que consiste en los elementos M. (…) Los fenómenos mentales son causados por los procesos que suceden en el cerebro en el nivel neuronal o modular, pero están realizados en el mismo sistema que consiste en neuronas organizadas en módulos… (…) Nada hay más común en la naturaleza que el que los rasgos de superficie de un fenómeno sean causados por micro-estructura y realizados en ella; y esas son, exactamente, las relaciones que son exhibidas por la relación de la mente con el cerebro. Las características intrínsecamente mentales del universo son las características físicas de alto nivel de los cerebros.” (Searle, 1995, pp. 432-433)

Un rasgo que Searle resalta de su concepción de intencionalidad es la direccionalidad:

> "La intencionalidad es aquella propiedad de muchos estados y eventos mentales en virtud de los cuales éstos se dirigen a, o son sobre o de, objetos y estados de cosas del mundo." (Searle, 1992, p. 18)

Según Salcedo Albarán (2004, p. 43), esto quiere decir, que para Searle algunos estados mentales tienen intencionalidad en la medida en que son acerca de algo; por ejemplo, las creencias, los temores y los deseos son intencionales en tanto que siempre debemos tener creencias, temores acerca de o sobre algo. No obstante, Salcedo Albarán señala que para Searle hay estados mentales como algunas formas de ansiedad o depresión que no son acerca de algo en especial. Es decir, que dichas formas de ansiedad, aunque son estados mentales, no son intencionales.

De acuerdo con Salcedo Albarán (2004, p. 43), en la concepción de intencionalidad de Searle también se debe tener en cuenta la distinción entre tener la *intencionalidad de* y la Intencionalidad. Así pues, tener la *intención de* hacer algo, es una forma más de Intencionalidad, de manera que no toda la Intencionalidad consiste en tener la *intención de.* Es decir, las intenciones no tienen un status especial dentro del conjunto de la Intencionalidad:

> "De esta manera, se puede asegurar que las intenciones no tienen ninguna importancia mayor dentro de la teoría de la Intencionalidad, que la que tendría cualquier otro tipo de estado Intencional, como las creencias, los deseos, los anhelos, etc. Así, según Searle, la intención de hacer algo es sólo una forma más de Intencionalidad; las intenciones son Intencionales, pero no abarcan

> todas las formas posibles de Intencionalidad." (Albarán, 2004, pág. 43)

De este modo, Searle denomina "Intencionalidad" la característica de direccionalidad de algunos estados mentales, e "intencionalidad al conjunto de *intenciones de* hacer algo (Albarán, 2004, pág. 43).

Salcedo Albarán señala (2004, p. 44), que para Searle la Intencionalidad no consiste en una relación común, como las relaciones que se dan cuando decimos que nos sentamos *sobre* una silla o caminamos *sobre* el piso:

> "En este último caso "sobre" una silla no es igual al "sobre" de tener una creencia "sobre" una persona. Esto, porque en la relación Intencional, puede no existir un objeto especifico o un estado de cosas del mundo acerca del cual se dirija la relación; por ejemplo, se puede tener la creencia de que el rey de Francia es calvo, aunque en la actualidad no exista un rey de Francia." (Albarán, 2004, pág. 44)

En lo que respecta a la Intencionalidad Searle hace una distinción entre Intencionalidad intrínseca e Intencionalidad derivada. Así, una Intencionalidad verdadera y legítima solamente se manifiesta, tal como afirma Salcedo Albarán (2004, p. 44) en un enunciado del tipo *"estoy sediento, muy sediento, porque no he bebido nada en todo el día"*, si este enunciado expresa una sensación verdadera de sed, entonces se implica el deseo de beber algo. Sin embargo, en un enunciado del tipo *"mi césped está sediento, muy sediento, porque no ha sido regado en una semana",* aunque se está usando el mismo término que en el enunciado anterior y se le está atribuyendo al césped la capacidad de tener sed, en realidad no se está

hablando del mismo tipo de sed. Pues el segundo enunciado sólo tiene, según Salcedo Albarán, propósitos metafóricos:

> "Mi césped, necesitando agua, estaría en una situación en la que yo estaría sediento, así que figurativamente hablando lo describo como si estuviera sediento." (Searle, 1996, p. 78)

Según Salcedo Albarán (2004, p. 44), en el primer enunciado hay una adscripción legítima de Intencionalidad intrínseca, porque si el enunciado es verdadero, entonces debe haber un estado Intencional en el objeto de adscripción. Mientras que el segundo enunciado, no consiste en algún tipo de Intencionalidad sino en un uso metafórico de tener sed. Es una adscripción del tipo *como si*, pero no se refiere a una característica intrínseca a algún sistema Intencional:

> "La Intencionalidad intrínseca es un fenómeno que los humanos y otros animales tienen, como parte de su naturaleza biológica (…). Es muy conveniente usar la jerga de Intencionalidad para hablar acerca de sistemas que no la tiene, pero que se comportan como si la tuvieran (…), pero es importante enfatizar que estas atribuciones son psicológicamente irrelevantes, porque no implican presencia de ningún fenómeno mental. La Intencionalidad descrita en estos casos es puramente como si." (Searle, 1996, p. 79)

Es precisamente este concepto de Intencionalidad intrínseca, que poseen los seres humanos como agentes biológicos, lo que nos diferencia de los dispositivos computacionales que defienden los teóricos de la IA fuerte, los cuales sólo poseen una Intencionalidad derivada. Esto es, pareciera que tuvieran rasgos de Intencionalidad, pero realmente no hay en dichos dispositivos fenómenos mentales que nos permitan hablar de Intencionalidad en el sentido biológico de la palabra, puesto que lo que hay es una adscripción del tipo *como si*.

Por tal razón, los dispositivos computacionales por más manipulación formal de símbolos que lleven a cabo por medio de reglas sintácticas, sólo poseen sintaxis y no semántica, pues la semántica es propia de los seres humanos como sistemas Intencionales. En este sentido, en los dispositivos que describen la IA fuerte sólo hay simulación de estados mentales, pero no una mente humana como tal. Sólo hay sintaxis, y la sintaxis como lo muestra el argumento de la habitación china, no es suficiente para la semántica.

Sin embargo, a esta idea de que los dispositivos computacionales sólo poseen sintaxis y no semántica, le han surgido varias réplicas adicionales a las que presenté en el primer capítulo de este trabajo[2], las cuales me ocuparé de presentar a continuación.

La primera, es la réplica del conexionismo. De acuerdo con Salcedo Albarán (2004, p. 60) esta réplica se encuentra en el artículo titulado *Could a machine think?* de Paul y Patricia Churchland. Esta réplica consiste en asegurar que los cerebros pueden ser computadores, pero computadores radicalmente diferentes a los concebidos por la IA clásica. Pues los dispositivos computacionales clásicos son digitales y están completamente programados por reglas pre-definidas sintácticamente, mientras que el cerebro no.

[2] La réplica del robot, la réplica de los sistemas, la réplica de la combinación y la réplica de las otras mentes.

Para los Churchland, las verdaderas dificultades que enfrenta la IA fuerte y la imposibilidad de que las máquinas tradicionales lleguen a constituir verdadera inteligencia y conciencia, se fundamenta en los fallos que resultan de no enfrentar el problema desde una interpretación de los sistemas nerviosos como máquinas en paralelo.

Además de esta dificultad, se encuentra tal como sostiene Salcedo Albarán (2004, p. 60) otra dificultad, a saber: que el funcionamiento de la neurona se encuentra enmarcado por un procesamiento analógico y no digital, en la medida que los índices de disparo neuronal varían constantemente en función de las entradas sensoriales.

Para los Churchland, Searle asegura que en esta interpretación de los sistemas nerviosos como máquinas en paralelo, también hay falencias semánticas. Para presentar su idea, Searle propone de acuerdo con Salcedo Albarán, que en lugar de pensar en una habitación china, se piense en un gimnasio chino, en el que muchas personas están organizadas en redes en paralelo ejecutando los mismos cómputos que ejecutaba la persona en la habitación china.

Frente a este argumento, los Churchland aseguran según Salcedo Albarán (2004, p. 60), que sería necesario contar con alrededor de 11 mil millones de personas porque un cerebro humano alberga alrededor de 10 mil millones de neuronas, cada una de ellas con un promedio de 1000 conexiones, lo cual representa la población de 1000 tierras. De este modo,

dicho modelo cósmico de personas procesando información podría representar un gran cerebro lento pero funcional. No obstante, los Churchland reconocen que la teoría del procesamiento en paralelo podría no ser correcta, sin embargo, tampoco hay a priori una garantía de que esto no constituiría pensamiento.

Salcedo Albarán señala (2004, p. 60) que Josef Moural en su artículo *The chinese room argument*, considera que Searle no reconoce el tipo de dispositivos computacionales que los Churchland tienen en mente, debido a que simplemente concibe la posibilidad de alterar la estructura original de las máquinas digitales tradicionales, imponiendo ejecuciones paralelas de varios algoritmos simultáneos, tal como sucede con la idea del gimnasio chino.

Sin embargo, esta propuesta de Searle es, de acuerdo con Moural, desafortunada, en la medida que sigue siendo posible que un solo hombre procese y obtenga el output total de la habitación china:

> "Uno podría llamar a este escenario Gripa China: todos los operadores del gimnasio excepto uno, se quedan en casa con la gripa, y el que se queda en el gimnasio tiene que hacer todo el trabajo hasta que los otros se recuperen." (Moural, 2003, p. 228)

Además, Moural señala que el argumento original de Searle no se aplica de manera idéntica a las posibilidades conexionistas, por el hecho de que también pueden ser simuladas por dispositivos computacionales:

> "(…) No es una parte del argumento de Searle que si algo se puede procesar en un computador como dato, debe carecer de semántica (piénsese, por ejemplo, en llamadas procesadas digitalmente.)." (Moural, 2003, pág. 228)

La segunda réplica, es la de la instanciación y la réplica de una mente más. En la medida que el argumento de Searle consiste en afirmar que se puede tener una instanciación de un programa computacional sin que ello implique la presencia de Intencionalidad, Moural presenta según Salcedo Albarán (2004, p. 61), tres maneras de refutar esta idea:

1) Negar la ausencia del estado mental requerido, 2) negar que el programa que Searle considera es correcto, y 3) negar que la configuración de la habitación china instancia el programa computacional.

En términos de Salcedo Albarán (2004, p. 61), Jerry Fodor pretende sustentar la tercera propuesta mostrando que la configuración que propone Searle no es, en estricto sentido, una máquina de Turing que instancia de manera correcta a un dispositivo computacional. Puesto que para Fodor, Searle usa un concepto débil de instanciación, basado únicamente en una correlación de la secuencia de estados entre el programa computacional y su instancia, en este caso el hombre en la máquina. Para Fodor, se requiere un concepto de instanciación en el que intervengan causas eficientes, las cuales según Salcedo Albarán (2004, p. 61) no quedan claramente especificadas por éste.

De esta manera, el argumento de Fodor deja intacto al propuesto por Searle en la habitación china, en la medida en que este último, a diferencia de Fodor sí ofrece un concepto de instanciación eficiente

especificado, en este caso, la persona que se encuentra dentro de la habitación manipulando los símbolos.

Salcedo Albarán afirma (2004, p. 61) que David Chalmers retoma la crítica de Fodor, pero con una versión mejorada de la réplica de los sistemas. Así pues, Chalmers asegura que los estados mentales de la persona que se encuentra dentro de la habitación son irrelevantes para la instanciación de un sistema computacional. Debido a que lo relevante para el sistema computacional son las relaciones dinámicas que se pueden encontrar entre los símbolos.

Para Chalmers, si se asume que lo verdaderamente importante es el entendimiento del hombre dentro de la habitación, entonces desde el principio de la argumentación de Searle se está asegurando que no se encontrará conciencia ni inteligencia en el sistema, en la medida que en el computador no hay nada que corresponda a la persona que manipula los símbolos en la habitación.

En este sentido, señala Salcedo Albarán (2004, p. 62), que de acuerdo con la idea de Chalmers, se puede decir que cuando Searle asegura que él en tanto sistema no entiende nada acerca de los símbolos chinos que procesa de acuerdo a reglas pre-definidas sintácticamente, en realidad está formulando una intuición y no está diciendo algo de lo que pueda estar seguro. Para Fodor no es el sistema quien concluye que no entiende nada, sino la instanciación propuesta por Searle, en tanto que, entre su

instanciación y el sistema computacional no hay nada que corresponda. Así pues, afirma Salcedo Albarán:

> "En esta medida Searle da un paso injustificado de una conclusión precedido por "me parece obvio que…" a una conclusión precedida por "yo no entiendo nada."" (Albarán, 2004, pág. 62)

La tercera réplica, es la de los conectivistas. De acuerdo con este enfoque computacional, es posible hacer evolucionar la IA fuerte de un modo cualitativamente distinto, que no consista en ofrecer más de lo mismo, aumentando la complejidad de los procesos computacionales. Esta nueva esperanza dentro del marco epistemológico de la teoría computacional de la mente tiene un nombre: *redes neurales.*

Navarro señala (2005, p. 269) que a diferencia de los programas de software convencionales, la redes neurales carecen de un código rígido pre-definido. Por el contrario, van generando su propio software a medida que interactúan con su entorno, mediante un proceso de aprendizaje prolongado en el tiempo.

Esta posibilidad parece marcar una diferencia cualitativa con los modelos computacionales de la IA fuerte atacados por Searle a principios de los años 80. Durante el proceso de aprendizaje con su entorno, el dispositivo computacional sería capaz de desarrollar su propia semántica y, por lo tanto, no limitarse a la aplicación sintáctica de reglas formales ya pre-definidas. De este modo, señala Navarro:

> "La intencionalidad estaría garantizada al no tratarse de un lenguaje impuesto sin referencia al exterior, sino de un lenguaje

> generado a través de la relación entre la máquina y el entorno circundante. Este tipo de máquinas son mucho más flexibles que las anteriores, y se adaptan a las circunstancias cambiantes del medio de un modo que para el ordenador convencional era imposible. Parece ciertamente que aprendan por ellas misma, que comprendan por ellas mismas las relaciones que se establecen entre las cosas." (Navarro, 2005, pág. 269)

A pesar del entusiasmo que pueda despertar este proyecto computacional de la mente, el argumento de la habitación china puede ser modificado de forma que atrape en sus redes no sólo a los ordenadores convencionales de la IA fuerte, sino también a las redes neurales recién estrenadas.

Para justificar este argumento Navarro recurre a una versión más sofisticada de la habitación china (2005, pp. 269-270): supongamos que el libro de instrucciones entregado a Searle era perfectamente correcto cuando él entro en la habitación. En aquel momento era posible simular a la perfección la conversación de cualquier hablante nativo del chino. Navarro nos pide que seamos lo suficientemente desconsiderados como para mantener a Searle encerrado en la habitación durante un largo periodo de tiempo.

Supongamos ahora que el idioma chino, como cualquier otro idioma vivo, va evolucionando con el paso de los tiempos, de modo que poco a poco el manual de instrucciones con el que Searle entró en la habitación ha ido quedando obsoleto. Así, hay ocasiones en las que sus respuestas ya no son válidas, y por tal razón, las personas fuera de la habitación empiezan a dudar

de la capacidad de comprensión del hombre de la habitación, en este caso Searle.

En ese momento, señala navarro (2005, p. 270), entran en juego los programadores del funcionamiento del invento. Viendo la desconfianza de las personas fuera de la habitación, deciden introducir una pequeña modificación en el diseño: cada vez que Searle responda utilizando una regla que ha quedado obsoleta, los programadores procederán a aplicarle una pequeña descarga eléctrica. En ese momento Searle pensará que ha debido realizar alguna operación equivocada, y tras repasar su manual comprobará que no es así, debido a que ha aplicado de manera correcta las reglas.

Según Navarro (2005, p. 270), ante el sufrimiento repetido por las descargas eléctricas, Searle supondrá que debe haber algo en el mundo externo que ha cambiado, haciendo que determinada regla sea obsoleta y, en consecuencia, inadecuada. Así pues, es probable que identifique de qué regla se trata en el momento en que recibe las descargas, y que vaya probando por ensayo y error otras alternativas. De este modo, Searle logra encontrar una regla cuya aplicación no implica descarga eléctrica, logrando con ello corregir el libro de reglas para adaptarlo al entorno, transformando el software original en función al medio cambiante con el que se relaciona.

Sin embargo, Navarro sostiene (2005, p. 270), que aun cuando Searle haya logrado imitar el proceso de aprendizaje de una red neural transformando su programa mediante una prolongada interacción con el

entorno, lo más probable es que siga sin entender los símbolos chinos y, por lo tanto, sin haber adquirido durante ese proceso semántica alguna. Pues la modificación de las reglas al entorno, no responde a una comprensión semántica de los símbolos, sino a una reacción constante de ensayo y error incitada por las descargas recibidas al aplicar reglas obsoletas.

Por lo tanto, el argumento de las redes neurales de los conectivistas habrá fracasado:

> "El código de software habría dejado de ser rígido: Searle habría imitado el proceso de aprendizaje de una red neural, transformando su programa durante una prolongada interacción con el medio. Pero, desgraciadamente, lo más probable es que siga sin comprender lo más mínimo acerca de a qué se refieren los términos. Puede que haya adquirido con el paso de los años esa inmensa paciencia que caracteriza a los artesanos chinos, pero no habrá adquirido en absoluto una semántica, y seguirá indefinidamente sin saber a qué se refieren los ideogramas que maneja… Parece por lo tanto que el argumento de la habitación china puede ser adaptado para hacer frente a las redes neurales que son, como dijimos antes, la nueva esperanza de los ordenadores para desarrollar una inteligencia artificial en sentido fuerte. Dado que los procesos de aprendizaje siguen desarrollándose según criterios sintácticos, no ofrecen un contacto con el mundo más auténtico que los programas convencionales. Nada les permite dar el salto de la sintaxis a la semántica: seguirán sin saber indefinidamente a qué se refieren sus signos." (Navarro, 2005, págs. 270-271)

Frente a estas réplicas elaboradas por los defensores de la IA fuerte, Searle desarrolla una nueva versión de su argumento[3]. Para ello, adicional a las tres premisas que constituyen el argumento original de la habitación china, a saber, 1) que los programas son enteramente sintácticos, 2) la mente tiene semántica, y 3) la sintaxis no es igual a, o por sí misma no

[3] . esta nueva versión del argumento de Searle ya fue presentada en el capítulo anterior del presente trabajo.

es suficiente para la semántica, por lo tanto, las mentes no son programas, Searle resalta la importancia del hecho de que la sintaxis no es intrínseca a ningún sistema físico y, en consecuencia, no está definida en términos físicos.

La argumentación desarrollada a partir de esta idea consiste en que la sintaxis es relativa al observador, y todas las características de la computación, en cuanto computación, son exclusivamente sintácticas. De esta manera afirma Salcedo Albarán:

> "Las personas cuentan con la capacidad de interpretar como símbolo cualquier objeto físico y esto manifiesta el hecho de que ningún símbolo es símbolo en virtud de que sea un objeto físico; los símbolos y la sintaxis son el resultado de la interpretación de los observadores y no de los objetos en sí mismos. Así pues, ningún patrón puede interpretarse como computación si no es porque alguien está asignando una interpretación de dicho patrón como computación." (Albarán, 2004, pág. 63)

En este sentido, vemos que para Searle la computación no es intrínseca a la física, sino que resulta de una interpretación hecha por el observador del sistema. Así, en el mundo se pueden encontrar objetos que son relativos al observador, y otros, que son independientes del observador. Pues hay rasgos del mundo que son intrínsecos y hay rasgos que son relativos, que dependen de un agente que observa:

> "Es, por ejemplo, un rasgo intrínseco del objeto que está enfrente a mí que tiene una determinada masa y una determinada composición química (…). Pero también se puede decir con verdad del mismo objeto que es un destornillador. Cuando lo describo como un destornillador, estoy determinando un rasgo del objeto que es relativo al observador o al usuario. Es un destornillador sólo porque la gente lo usa como (o lo ha hecho para el propósito de servir como, o lo ve como) un destornillador.(Searle, 1997, p. 29)

De este modo, para Searle (1996, p. 215) la caracterización de un proceso como computacional es una caracterización de un sistema físico desde fuera, y la identificación del proceso no identifica un rasgo intrínseco de la física, es esencialmente una caracterización relativa al observador. Por tal razón, si la computación se define en términos de la asignación de sintaxis, entonces todo puede ser un ordenador digital, puesto que a cualquier cosa se le podrían hacer adscripciones sintácticas. Es decir, se podría interpretar cualquier cosa en términos computacionales:

> "Así pues, surge la pregunta: ¿Es la computación un rasgo intrínseco de un sistema físico como el cerebro, o es un rasgo que resulta de una interpretación del cerebro como un computador? Según Searle, hay casos limitados en que, en estricto sentido, los seres humanos computan, "por ejemplo, ellos computan la suma de 2+2 y obtienen 4." En estos casos, se puede decir que la computación es independiente del observador "en el sentido en que no se requiere un observador externo que los interprete como computando para que ellos estén computados." Ahora bien, ¿Qué sucede con los computadores comerciales que todos conocemos? ¿Qué parte de la física o la química de sus impulsos eléctricos permiten la constitución de símbolos? Según Searle, no hay nada, ni físico ni químico, que convierte a dichos impulsos en símbolos porque símbolos o sintaxis no se refiere a un rasgo intrínseco como electrón o placa tectónica. Así pues, la siguiente es la respuesta de Searle a la pregunta ¿El cerebro es intrínsecamente un computador digital? No, porque algo es un computador solamente relativo a las asignaciones de una interpretación computacional." (Albarán, 2004, págs. 63-64)

Esto no quiere decir que no sea posible asignar interpretaciones computacionales al cerebro, pues de hecho, se le puede asignar tal como hemos visto, interpretaciones computacionales a cualquier proceso. Para Searle una consecuencia principal de este argumento es que:

> "No se puede descubrir procesos computacionales en la naturaleza, independientemente de la interpretación humana porque cualquier procesos físico que se pueda encontrar es

> computacional solamente relativo a alguna interpretación.” (Searle, 1997, p. 16)

En este sentido, podemos ver que para Searle la computación no es un rasgo intrínseco al cerebro como piensa la IA fuerte, sino el resultado de una interpretación por parte de un agente observador, quien adscribe al cerebro humano rasgos computacionales. En otras palabras, la computación no es un proceso de esta máquina orgánica que es el cerebro humano, como si lo es el índice de disparo neuronal.

De acuerdo con Salcedo Albarán (2004, p. 64), para Searle el presente argumento es mucho más profundo que el inicialmente planteado en la habitación china, porque aparte de mostrar que la semántica no es intrínseca a la sintaxis, muestra que la sintaxis no es intrínseca a la física. Por lo tanto, los argumentos en los que se apoya la IA fuerte para sustentar su tesis computacional se vienen abajo.

Según Salcedo Albarán (2004, p. 65), el argumento de Searle se dirige en primera instancia, a resaltar los poderes causales del cerebro, en cuanto a su capacidad de constituir intencionalidad y, en segunda instancia, a señalar que dichos poderes causales no están presentes en un programa por sí solo. Pues Searle siempre señala que la premisa fundamental de la IA fuerte es la idea de que una mente puede constituirse a partir de la instanciación de un programa y, por lo tanto, el soporte físico sobre el que se ejecute el programa es irrelevante, pues el programa por sí solo cuenta con los poderes causales necesarios para constituir mentes.

Para Salcedo Albarán (2004, p. 65), Searle no pretende refutar la idea de que un dispositivo computacional puede causar estados mentales, sino la idea de que el programa por sí solo puede hacerlo. Searle reconoce que puede haber máquinas que causen mentes, pero solamente máquinas que tengan una configuración idéntica a la cerebral, porque en estricto sentido, la única máquina que cuenta con los poderes causales necesarios para construir estados mentales, es la máquina orgánica que conocemos como cerebro:

> "Es importante ver lo que es afirmado y lo que no es afirmado por mi argumento. Supóngase que planteamos la pregunta que mencioné al principio. ¿Puede pensar una máquina? Bien, en algún sentido, desde luego, todos somos máquinas. Podemos interpretar la materia que tenemos dentro de nuestras cabezas como una máquina de carne. Y desde luego podemos pensarlo todo. Así, en un sentido de "máquina", a saber: ese sentido en el que máquina es solamente un sistema físico que es capaz de realizar cierto género de operaciones, en ese sentido, todos somos máquinas, y podemos pensar. Así, trivialmente, hay máquinas que pueden pensar." (Searle, 1985, p. 41)

Además, para Salcedo Albarán (2004, p. 65), si nos alejamos del soporte físico y suponemos la relevancia causal de únicamente los programas, como según Searle lo hace la IA fuerte, caeremos en un error, debido a que los programas sin implementar son abstractos, sistemas formales que no tienen poderes causales de ningún tipo y, como tales, no pueden producir mentes. Así, el proyecto computacional de la mente que pretende la IA fuerte termina fracasando.

A pesar de estos planteamientos por parte de Searle para justificar su argumento de la habitación china y, derrumbar con ello, cualquier

hipótesis de la IA fuerte que intente fundamentar computacionalmente a la conciencia, se han elaborado algunas críticas actuales por parte de algunos defensores de la IA fuerte, que pretenden mostrar las falencias conceptuales que aun subyacen en el argumento de la habitación china, y de las que el argumento de Searle parece no escapar.

La primera, es la imposibilidad del supuesto de la sustitución en Searle. Este supuesto consiste, en que todo el complejo de cables, conexiones y placas metálicas de las que está compuesto un sistema computacional es sustituido por un ser humano. Pues para Searle podemos sustituir el sistema, todo lo que él es, por una persona que realice las mismas tareas que le asignamos a aquel.

De acuerdo con Artiles Arbelo (2001, p. 86), esta sustitución es imposible y, por tal razón, no podemos probar si la teoría de la mente en cuestión (la de la IA fuerte) es errónea bajo el supuesto de la sustitución. Para Artiles Arbelo es importante señalar en qué consiste el argumento final de la habitación china de Searle para los propósitos de la crítica. El argumento de Searle es que la mera manipulación sintáctica de símbolos no es suficiente para la comprensión de un lenguaje, en este caso, el chino. Necesitamos de la semántica y ésta no es posible si sólo nos remitimos a dicha manipulación de símbolos.

Una vez señalado el argumento final de Searle, Artiles Arbelo utiliza un diseño semejante al propuesto por Searle para demostrar la imposibilidad

del supuesto de la sustitución (2001, p. 91). El diseño consiste en que dentro de una habitación se halla un hombre (Searle) que hará las veces del conjunto de cables, conexiones y placas metálicas constituyentes de un dispositivo computacional. Se supone además que Searle posee un dominio perfecto del castellano en el mismo sentido en el que se lo suponemos a un hablante nativo del castellano. El primer libro de reglas del argumento de Searle está escrito en castellano, así como el segundo libro que se le suministrará más adelante.

Además de suponer en Searle un dominio perfecto del castellano, podemos suponer también la capacidad sintáctica de manipulación formal y de interpretación de la serie de símbolos que aparecen en cada uno de los dos libros de reglas. De esta manera Searle posee las dos características necesarias para el manejo de un lenguaje: sintaxis y semántica.

Así, si queremos hacerle alguna pregunta en castellano nos respondería de manera parecida a cualquier hablante nativo del castellano. Ahora bien, se le pide que coteje los símbolos chinos procedentes de una de las cestas con los símbolos de la otra, según el primer libro de reglas que se le ha dado al comienzo; igualmente, mediante el segundo libro de reglas que se le ha entregado ha de enviar a modo de respuestas, una serie de símbolos chinos como respuesta a cada pregunta que le introducimos por medio de una ranura de la habitación.

En todo este proceso Searle sostiene que lo único que hace es manipular símbolos chinos y ni siquiera sospecha que se trata de un juego de preguntas y respuestas. Según Artiles Arbelo (2001, p. 91), hasta este punto todo coincide con el argumento desarrollado por Searle en el argumento de la habitación china, pues no se ha introducido nada nuevo.

Llegados a este punto Artiles Arbelo introduce el concepto de "ser neutral respecto a la semántica" (2001, p. 91), refiriéndose con ello a la suspensión de la adscripción de significado a los símbolos manipulados formalmente. Esto con el propósito de dar lugar a la cuestión de si Searle es o no neutral respecto a la semántica que ya poseía como hablante nativo del castellano.

La respuesta que nos ofrece Artiles Arbelo es que Searle no puede ser neutral respecto a la semántica, debido a que hace uso constante de la semántica anterior como consecuencia de ser un hablante nativo del castellano. Por lo que la posibilidad de ser neutral respecto a la semántica no puede darse nunca, puesto que esto incapacitaría al propio Searle para entender su propio lenguaje:

> "No podemos hacer que Searle se dedique sólo a manipular símbolos, no podemos amputar sus capacidad semántica a riesgo de que no comprenda su propio lenguaje del que es nativo. Afortunadamente la sintaxis y la semántica se encuentran revueltas, de alguna forma, en un hablante nativo de cualquier lenguaje y por ello, afortunadamente también, no podemos prescindir de una de esas capacidades, en particular la semántica, en Searle. Mantengámos entonces esa capacidad semántica en el profesor Searle y propongámos que éste haga más que manipular símbolos exclusivamente. Pero esto quiere decir, en contra de nuestro supuesto de sustitución, que no se comportará como lo

> haría un computador digital, pues éste sólo hace manipulación de símbolos formales. Searle debe hacer algo más que eso antes de que pueda dar una sola contestación a nuestras preguntas. Por lo tanto, la primera conclusión a la que llegamos es que el profesor Searle, aunque le pese, no puede hacer sólo lo que hace un computador digital y, por ello, no puede sustituirlo en la habitación china." (Arbelo, 2001, pág. 92)

De este modo podemos ver, que el que Searle no pueda ser neutral respecto a la semántica del castellano como hablante nativo de éste, mientras coteja los símbolos chinos en la habitación, imposibilita el supuesto de la sustitución que pretende el propio Searle en su argumento. En la medida, que Searle al hacer uso constante del dominio semántico del castellano, haría más que manejar simples símbolos y, en consecuencia, su comportamiento mientras está en la habitación sería diferente al de un dispositivo computacional, por lo que no podría sustituirlo. Pues éste último, sólo hace mera manipulación formal de símbolos mediante reglas sintácticamente pre-definidas.

Artiles Arbelo señala (2001, p. 92), que el hecho de que Searle no pueda ser neutral respecto a la semántica del castellano, demuestra también lo contrario a lo expuesto por él en su argumento de la habitación china, a saber, que no puede haber una semántica de unos símbolos chinos a partir de un hablante nativo del castellano que ya posee, por definición, la capacidad semántica y sintáctica de un lenguaje, con la mera manipulación de esos símbolos chinos.

Así, afirma Artiles Arbelo (2001, p. 92), que si poseemos un lenguaje con respecto al que estamos capacitados sintáctica y semánticamente,

significa esto que tenemos una interpretación dada sobre los símbolos de nuestro lenguaje, y que podemos crear a partir de esa interpretación, una interpretación que englobe cualquier conjunto de símbolos que añadamos a los de nuestro lenguaje.

Ahora bien, se podría objetar que la interpretación que hemos dado a esos símbolos a partir de nuestro dominio sintáctico y semántico de nuestro lenguaje, no sea la correcta. Es decir, la interpretación estándar que haría cualquier hablante nativo del chino. Pero lo crucial del argumento que propone Artiles Arbelo, es que se ha conseguido que el hombre en la habitación posea al menos una interpretación, lo que basta para refutar el argumento de Searle.

En otras palabras, no se trata de si la interpretación que hemos dado a esos símbolos chinos sea o no la adecuada, sino que sobre el manejo sintáctico de esos símbolos, se ha elaborado una interpretación de tales símbolos, y eso es semántica. Por lo que el enunciado de Searle se hace refutable y, en consecuencia, el supuesto de la sustitución que pretende en su argumento se hace imposible. Pues a diferencia de cualquier sistema computacional, Searle estaría capacitado semánticamente para dar una interpretación a los símbolos que maneja sintácticamente, debido a que no podría ser nunca neutral respecto a la semántica del castellano como hablante nativo del mismo, haciendo con ello uso constante de su lado semántico mientras está en la habitación:

> “La interpretación que podemos hacer puede no ser la que haría un hablante nativo del chino pero, si recordamos, Searle dice que es imposible cualquier semántica del chino sólo a partir de la sintaxis de un conjunto de símbolos realizada por un hablante competente de otro lenguaje. Pero nosotros hemos conseguido que nuestro hombre en la máquina posea al menos una interpretación, lo que nos basta para refutar el enunciado de Searle.
>
> Así las cosas, resulta que Searle no puede refutar la Inteligencia Artificial fuerte a través del argumento de la habitación china. No puede hacerlo al fallar el supuesto de la sustitución. Y ello, por dos razones: a) no sólo manipulamos símbolos, estamos además capacitados semánticamente para su interpretación y no podemos separar nuestros respectivos lados sintácticos y semánticos, y b) en base a esa semántica previa que nos define como hablantes nativos en sentido usual, podemos crear una interpretación sobre cualquier conjunto de símbolos que nos sean dados, en particular, un conjunto de símbolos chinos.” (Arbelo, 2001, pág. 93)

De esta manera, si Searle pretendía rebatir la posibilidad de sustitución de un hombre, y todos sus estados y procesos mentales, por un dispositivo computacional con sus programas, ha conseguido el efecto contrario, pues como se señala Artiles Arbelo en la cita anterior, ha demostrado indirectamente que la sustitución de ese mecanismo computacional por un ser humano no es posible.

Según Artiles Arbelo (2001, p. 93), si es posible que hagamos sólo y exclusivamente manipulación formal de símbolos no seremos capaces ni siquiera de duplicar el comportamiento de un dispositivo computacional:

> “Parce irónico que Searle se esfuerce en diferenciar la simulación y la duplicación de los poderes causales del cerebro por parte de un computador, cuando nosotros somos estrictamente incapaces de duplicar exclusivamente su comportamiento. No somos capaces de ponernos en el lugar del computador sin hacer nada más. El presupuesto de la sustitución falla.” (Arbelo, 2001, pág. 93)

Además, tal como sostiene Artiles Arbelo (2001, p. 94), si se aceptan los comentarios más arriba expuestos en torno a la imposibilidad de la sustitución, el hombre en la habitación (Searle) tendría la posibilidad de adquirir conocimientos sobre lo que ocurre más allá de las paredes de la habitación china. En este sentido, si es cierto que Searle, al entrar en la habitación continua haciendo uso de su lado semántico, que por definición le corresponde como hablante nativo del castellano, entonces podríamos también adscribirle la intencionalidad asociada que Searle adscribe a un hablante nativo.

De modo que con toda esa capacidad, no sólo podría responder a todo lo que se le pregunte, sino también formular preguntas diferenciándose con ello de un dispositivo computacional cualquiera. Obteniendo así, tanta información sobre lo que ocurre fuera de la habitación china que estaría en condiciones de formular teorías sobre el modo en que ocurren las cosas fuera de la habitación. Así pues, afirma Artiles Arbelo:

> "Aquella pequeña ranura por donde se le enviaban los libros de reglas y los conjuntos de símbolos se iría transformando, así, en una gran ventana hacia el exterior. Su comprensión no se limitaría a la que pudiera tener sobre unos símbolos abstractos introducidos en la sala. Iría mucho más allá y la habitación china, lejos de ser un receptáculo hermético, se transformaría poco a poco en una especie de invernadero acristalado. Ahora la habitación china tendría ventanas y, dentro, el profesor Searle podría disfrutar de un paisaje que él mismo pensaba no vería jamás." (Arbelo, 2001, pág. 94)

La segunda crítica al argumento de Searle, tiene que ver con el hecho de que el argumento de la habitación china tal y como está diseñado excluye al resto de los seres humanos de la categoría de sujetos mentales

intencionales. En la medida, que el criterio de asignación de mentalidad es tan rígido que sólo lo puede superar uno mismo. En otras palabras, tal como señala Navarro (2005, p. 267), en caso de asumir que la perspectiva de primera persona es infalible, y ésta es una condición indispensable del argumento de Searle, el problema de las otras mentes se hace insoluble. De aquí, que el argumento de la habitación china, tal como está formulado, nos conduce a un solipsismo.

Para justificar su argumento Navarro nos invita a que analicemos el argumento de Searle desde otra perspectiva, para ello apela a la utilización de dos lenguajes distintos: el inglés y el chino (2005, pp. 271-272). El argumento consiste en que las instrucciones que recibe Searle están en inglés; la historia que se le ofrece a través de la ranura de la habitación, así como cada una de las preguntas que se le formulan con respecto a ella, están en chino.

Navarro nos pide que hagamos la suposición de que nosotros somos capaces de manejar ambos idiomas y que, por lo tanto, mantenemos una conversación diferente en cada uno de ellos.

En primer lugar, mantenemos una conversación por escrito en chino, a través de las ranuras de la habitación. Para no establecer ningún prejuicio acerca de la identidad de la persona con la que mantenemos esta primera conversación, navarro nos invita a que la llamemos X. A continuación, mantenemos una segunda conversación, esta vez oral y en

inglés con Searle en persona. Hay, por lo tanto, dos conversaciones: una con X, por escrito en chino, y otra con Searle, oralmente en inglés.

Ahora bien, si preguntamos a ambos sujetos si comprenden el chino, evidentemente surgirá una contradicción. Puesto que X pensará, como afirma Navarro (2005, p. 272), que no se les estará tomando en serio y, en consecuencia, se indignará frente a nuestra desconfianza de su comprensión del chino, debido a que todo este tiempo se ha estado conversando en dicho idioma. Mientras que Searle responderá que no entiende ni una sola palabra en chino y que, por lo tanto, la escritura china no le resulta más que un montón de garabatos sin sentido.

Según Navarro (2005, p. 272), el peso de todo el argumento reside en que, para Searle, debemos concederle mayor credibilidad a su discurso en inglés que al emitido por X en chino, lo cual nos conduce a la cuestión fundamental: ¿Cuál es el criterio con el que podemos otorgar mayor credibilidad al discurso de uno sobre el discurso del otro?

La respuesta de Searle a la pregunta es: el cerebro. Aquí está la diferencia, pues la conciencia es una cualidad del cerebro, que es producida causalmente por procesos neurobiológicos llevados a cabo en él y que nos brinda la capacidad de comprensión. Mientras que los chips de los sistemas computacionales carecen de esa capacidad causal. Sólo imitan el resultado de la inteligencia humana, pero no son capaces de generar una conciencia similar a la nuestra y esta diferencia es, según Searle, una cuestión empírica.

De modo que si se trata de una cuestión empírica, entonces no habría más que mostrar el dato empírico que demuestra que un cerebro es consciente, mientras que un dispositivo computacional no lo es.

De acuerdo con Navarro (2005, p. 273), aquí empiezan los problemas del argumento de Searle que nos arrastran al solipsismo. Pues como afirma Navarro, para Searle la subjetividad es un hecho objetivo desde el punto de vista biológico, en la medida que es una cualidad de la conciencia que emerge de los procesos neurobiológicos que se producen en el cerebro como sistema biológico. Acepta además, sin tapujo alguno, que dicha cualidad de la conciencia es un hecho objetivo que carece de criterios de tercera persona que lo corroboren. Puesto que sólo se puede tener acceso a ésta en una epistemología de la primera persona.

Así pues, sostiene Navarro (2005, p. 273), que más allá de las aseveraciones subjetivas de Searle, expresadas en el lenguaje, y de sus afectos en la conducta, carecemos de una prueba empírica de la existencia de la conciencia intencional como prueba de que el cerebro de Searle realmente lleva a cabo un proceso de comprensión semántico de los símbolos que maneja dentro de la habitación. Abocándonos con ello a una epistemología de la primera persona, y cuya esperanza de sostener el argumento sería apelando a la experiencia subjetiva. En este sentido, sostiene Navarro:

> “Si no tenemos más datos para certificar la apariencia de conciencia intencional (…) ¿Cómo sabremos que el dato físico en

> cuestión está correlacionado con la conciencia intencional intrínseca, y no con la mera computación? La articulación de las perspectivas objetiva y subjetiva, de la tercera y la primera persona, parece imposible por mucho que Searle crea haberlo conseguido. Su esfuerzo por lograr esa síntesis es loable, pues se enfrenta a la cuestión en lugar de dejarla de lado, como ocurrió con el conductismo o con gran parte del funcionalismo. Pero no parece en realidad que su teoría de la conciencia escape a las consecuencias de una epistemología de primera persona, pues la apelación a la experiencia subjetiva sería el único apoyo que queda para mantener el argumento de la habitación china. De ahí que la única posibilidad de salvar el argumento sea despejar la incógnita X." (Navarro, 2005, págs. 273-274)

Al apelar a la experiencia subjetiva, Searle señala que él mismo es X, es decir, es una misma conciencia la que sostiene las dos conversaciones. No se trata de dos mentes enfrentadas, sino de una misma mente que en chino no sabe lo que dice y en inglés sí. Esta apelación a la experiencia de la primera persona, salvaría según Navarro (2005, p. 274) el argumento de no ser porque nadie ha sostenido nunca que Searle, en tanto Searle, comprenda chino en este experimento. Además, no se trata ya de si Searle comprende o no el chino, sino de si hay una propiedad emergente en el sistema cerebral que permita a X comprender chino, esto es, una conciencia intencional que pueda ser demostrada.

Navarro nos resume su idea afirmando que el argumento de Searle para que le creamos a él y no a X es que, según él, X no sabe a qué se refieren los signos que maneja, mientras que Searle si posee una intencionalidad gracias a la cual comprende el lenguaje que utiliza (2005, p. 274). Pero desgraciadamente, tal como hemos señalado, Searle no tiene ni puede tener modo alguno de mostrarnos a nosotros la existencia de su propia conciencia intencional, pues no contamos con los criterios objetivos

necesarios para poder demostrar en qué momento hace aparición la conciencia intencional en el cerebro de Searle y, por consiguiente, darle mayor credibilidad al discurso de él y no al de X. De este modo, afirma Navarro:

> "En realidad, el problema que encuentra Searle a la hora de atribuir intencionalidad y conciencia a X es el mismo que podemos encontrar nosotros a la hora de atribuir al propio Searle esa intencionalidad y esa conciencia intrínsecas que dice poseer. No se trata ya de mostrar si los ordenadores tienen o no una mente intencional. La habitación china nos introduce en un atolladero mucho más preocupante: apelando a la infalibilidad de la primera persona, el argumento nos obliga a encerrarnos en el solipsismo, por mucho que Searle quiera escapar de esa posición." (Navarro, 2005, pág. 274)

Así pues, el argumento de Searle nos conduce de acuerdo con Navarro, a un problema que ha perseguido como sombra a la filosofía del siglo XX cada vez que se ha ocupado de la subjetividad: el problema de las otras mentes. Para Navarro (2005, p. 275), sólo hay una manera de salir del callejón sin salida en el que nos ha metido Searle, y es redefiniendo nuestros términos fundamentales como conciencia e intencionalidad intrínseca, de modo que sean epistemológicamente accesibles. Sólo así podremos demostrar si en un sistema, ya sea biológico o computacional, se está produciendo conciencia intencional. Pues la manera como están definidos dichos conceptos, nos encierra en una epistemología de primera persona de la cual es difícil salir, debido a que carecemos de criterios objetivos para demostrar la existencia de una conciencia intencional, en este caso la de Searle:

> “Dado que dependen de una epistemología de la primera persona, carecemos por principio de un criterio de tercera persona para demostrar su presencia. Suponer que esa situación es meramente provisional es sólo un modo de enmascarar el problema. Por el contrario, la única salida a la habitación china pasa por una profunda reformulación de nuestros conceptos de conciencia e intencionalidad, y por una reflexión determinada acerca de qué entendemos como perspectivas de primera y tercera persona.” (Navarro, 2005, pág. 275)

En el capítulo que sigue me ocuparé de considerar las ideas analizadas hasta ahora, con el propósito de mostrar cuáles de ellas son más razonables. En otras palabras, trataré de resumir los principales argumentos de ambos enfoques teóricos y, exponer a partir de tal análisis, el punto de vista que considere que nos ofrece una mejor explicación del funcionamiento de la vida mental de los seres humanos.

Capítulo IV

Consideraciones finales: ¿Teoría computacional o enfoque neurobiológico de la mente?

El problema de la mente humana ha sido analizado a la luz de dos puntos de vista diferentes e inconciliables, puesto que la afirmación de una de estas posturas excluye a la otra. Por un lado tenemos la tesis de la IA fuerte, para la cual, el cerebro es simplemente un computador digital y la mente es solamente un programa de computador. Es decir, la mente es al cerebro lo que el programa es al hardware del computador. Pues el cerebro es uno de un número indefinidamente extenso de diferentes géneros de hardware de computador que podrían servir de soporte a los programas formales que constituyen la inteligencia humana.

Según esta tesis computacional, cualquier sistema físico que tuviese el programa correcto con los *inputs* y los *outputs* correctos tendría una mente en exactamente el mismo sentido que los seres humanos. En otras palabras, al funcionamiento normal de cualquier dispositivo computacional se le puede atribuir cualidades mentales, incluso a los dispositivos mecánicos más simples como un termostato.

Pues cualquier sistema que sea capaz de manipular símbolos de una manera correcta es capaz de inteligencia en el mismo sentido que la inteligencia humana. Los procesos mentales son procesos computacionalmente definidos.

Por otro lado, tenemos la tesis neurobiológica de la mente que defiende Searle. Para Searle (1996, pp. 101-102), una idea básica en nuestra concepción del mundo es que los seres humanos y otros animales superiores, son parte del orden biológico como cualquier otro organismo. Los seres humanos son una continuación del resto de la naturaleza.

De esta manera, las características biológicamente específicas de estos animales, tales como la posesión de un sistema rico de conciencia y mayor inteligencia, su capacidad para el lenguaje y para el pensamiento racional, son fenómenos biológicos como cualquier otro fenómeno biológico. Son todos estos, según Searle, rasgos del fenotipo. Son el resultado de la evolución biológica como cualquier otro fenotipo.

En otras palabras, la conciencia es un rasgo biológico de los cerebros humanos y de ciertos animales. Está causada por procesos neurobiológicos, y es una parte del orden biológico natural, tales como la fotosíntesis, la digestión o la mitosis.

Así, los productos de la evolución biológica y los organismos están hechos, en términos de Searle, de subsistemas denominados células. Algunos de estos organismos desarrollan subsistemas de células nerviosas

que concebimos como sistemas nerviosos, algunos de los cuales son capaces de causar y mantener procesos y estados subjetivos conscientes.

Dichos estados y procesos mentales conscientes tienen un rasgo especial que no poseen otros fenómenos biológicos o naturales: la subjetividad. Lo subjetivo tal como lo define Searle, se refiere a una categoría ontológica de la conciencia, no a un modo epistemológico.

Así, cuando decimos “tengo un dolor de cabeza”, el enunciado es objetivo, pues su verdad depende de la existencia de un hecho real y no de actitudes u opiniones de los observadores. Pero el fenómeno del dolor mismo, tiene un modo subjetivo de existencia y es en este sentido en el que la conciencia es, de acuerdo con Searle, subjetiva. Pues para que un dolor sea un dolor, debe ser un dolor de alguien.

A pesar de las diferentes dificultades que presentan ambos enfoques, considero que el enfoque que mejor nos ofrece una explicación de la mente humana, es el propuesto por Searle. Pues según Searle, lo que realmente simulan los ordenadores es la sintaxis del cerebro, no su semántica. Por lo tanto, lo que hacen los computadores o los robots diseñados por la IA fuerte es una mera simulación de los sistemas formales, pero no de la inteligencia humana ni tampoco de la intencionalidad propiamente dicha.

En otras palabras, no hay en tales diseños computacionales una intencionalidad intrínseca, sino derivada. Pues por más que parezca que una

máquina para jugar ajedrez posea racionalidad práctica a la hora de llevar a cabo los movimientos de cada una de las piezas del juego y que, en consecuencia, pueda aparentar actuar sobre la base de ciertas creencias, deseos, preferencias o cualesquiera otro fenómeno intencional, no hay en su accionar nada de eso. Lo que hay es una simulación de procesos mentales, que responden a un programa formalmente sintáctico.

En este sentido, una máquina no puede experimentar el hambre, la sed, el dolor, la alegría, la tristeza, el deseo, la esperanza, la angustia, el temor, la creencia sobre algo. No pueden actuar intencionalmente, su conducta se encuentra determinada por un programa meramente formal. Es decir, no se explica en términos de creencias, deseos, esperanzas etc., sino en términos de software.

De esto se desprenden dos consecuencias. La primera, que un sistema computacional, al no poder experimentar dolor o cualquier otro fenómeno intencional de la conciencia, queda privado de la experiencia de primera persona. Como por ejemplo la experiencia de saborear una cerveza, la experiencia de escuchar la Novena Sinfonía de Beethoven, la experiencia de oler una rosa, de ver el atardecer o la experiencia de sentir dolor (Searle, 2007, p. 6). Un sistema computacional está imposibilitado para tener experiencias subjetivas de estados mentales conscientes.

La segunda, es que no podemos imputarle responsabilidad moral a una máquina por su comportamiento, en la medida que su conducta no es

intencional y voluntaria, sino mecánica (responde a programas formales). Nosotros los seres humanos, a diferencia de las máquinas, actuamos bajo la base de nuestras creencias, deseos y demás fenómenos intencionales. Por lo que nuestras acciones son producto de nuestras decisiones y, por lo tanto, se justifican sobre la base de motivos y razones. En un agente en condiciones normales se ejerce con libertad, de modo que es responsable moralmente de su proceder.

En otras palabras, el que actuemos consciente y voluntariamente y podamos dar razones sobre nuestra conducta nos hace, a diferencia de los sistemas computacionales, responsable de nuestras decisiones y acciones.

La propuesta neurobiológica de Searle nos ofrece una mejor explicación de la conciencia. No trata la conciencia como fenómeno oculto e inaccesible a la ciencia, sino como un fenómeno biológico que puede ser estudiado científicamente como cualquier otro fenómeno biológico del mundo. Si podemos estudiar científicamente otros fenómenos biológicos, entonces ¿Por qué no poder estudiar también la conciencia, si ésta es el producto de procesos neurobiológicos llevados a cabo en el cerebro y su estructura?

Sería absurdo aislar del estudio del cerebro el estudio de la conciencia, así como sería absurdo, según Searle, estudiar el estómago sin estudiar la digestión o la genética sin estudiar la herencia de caracteres (2007, p. 1). Puesto que así como la digestión es una macro-propiedad del

estómago causada por una micro-propiedad del mismo como lo es su sistema digestivo, de la misma manera, hay que considerar los estados mentales como macro-propiedades del cerebro producidas por ciertas micro-propiedades del mismo (su estructura neuronal). Por lo que el estudio del cerebro, sus funciones y procesos, debería llevar aparejado el estudio de la conciencia humana.

La perspectiva neurobiológica de la conciencia explica el fenómeno de los estados consciente a nivel cerebral y neuronal. De acuerdo con esta postura, la conciencia emerge biológicamente del sistema o la estructura cerebral. Es decir, necesitamos de un cerebro para que se puedan dar en nosotros estados conscientes o neurológicos. Debido a que es precisamente en la estructura del cerebro donde se llevan a cabo los procesos neurobiológicos que causan nuestros estados subjetivos o conscientes.

Si bien es cierto que aún desconocemos qué son exactamente los correlatos neurobiológicos de la conciencia y cuáles de esos correlatos son en realidad causalmente responsables de la producción de la conciencia, en un futuro, con los avances de la neurociencia, tenemos la expectativa de que se puede llegar a iluminar el misterio de la conciencia.

A medida que avanza la ciencia vamos descubriendo cosas nuevas acerca del mundo, del ser humano y de ese órgano biológico tan complejo como lo es el cerebro humano, y sus manifestaciones conscientes e intencionales.

Antes desconocíamos el ADN, hoy día con los avances de la ciencia sabemos qué es un ácido nucleico que contiene instrucciones genéticas usadas en el desarrollo y funcionamiento de todos los organismos conocidos y algunos virus, y es responsable de su transmisión hereditaria.

En otras palabras, el estudio de la genética nos ha permitido comprender qué es lo que ocurre en el ciclo celular y reproducción celular de los seres vivos, y cómo puede ser posible que entre seres humanos se transmitan características biológicas y fenotípicas de apariencia y personalidad.

Igualmente hemos podido comprender gracias al desarrollo de la ciencia, el funcionamiento de otros fenómenos biológicos tanto de la naturaleza como de nuestro propio organismo. Actualmente conocemos muchos de los fenómenos biológicos que se llevan a cabo en el cerebro. Por ejemplo, sabemos que la memoria es el resultado de las conexiones sinápticas repetitivas entre las neuronas, lo que crea redes neuronales o potenciación a largo plazo (intensificación duradera en la transmisión de señales entre dos neuronas que resultan de la estimulación sincrónica de ambas).

Conocemos también muchas de las funciones complejas del cerebro, tales como las funciones del hemisferio cerebral, el hemisferio derecho, el lóbulo occipital, el lóbulo temporal, el lóbulo frontal, el hipocampo, el sistema

límbico etc., la neurociencia nos ha suministrado muchísimos conocimientos acerca del funcionamiento del cerebro humano.

Sin embargo, las áreas cerebrales que gobiernan algunos fenómenos como la memoria, el pensamiento, las emociones, la personalidad, la conciencia y otros fenómenos intencionales, son difíciles de localizar. Es decir, aún desconocemos su ubicación precisa en el cerebro. No obstante, se guarda la esperanza de que con el desarrollo de la neurociencia se pueda dar respuesta a tales inquietudes.

La conciencia no es un fenómeno oculto e inaccesible a la ciencia, por el contrario, es un fenómeno biológico, que puede ser estudiado desde un punto de vista científico en el mismo sentido en el que se estudia la genética, la digestión y la fotosíntesis.

Así pues, el estudio neurocientifico de la conciencia junto con el análisis filosófico, no sólo podrían ayudarnos a clarificar nuestros conceptos de conciencia e intencionalidad, sino también a resolver el misterio de la conciencia, puesto que lo que actúa en la causación de la conciencia y sus manifestaciones intencionales no son procesos computacionales, sino biológicos como la unidad neuronal y los procesos llevados a cabos en la estructura cerebral. Todos ellos, fenómenos completamente biológicos y no computacionales.

En otras palabras, el estudio de la conciencia como fenómeno biológico causado por procesos neurobiológicos del cerebro, y el análisis

filosófico de los conceptos de conciencia e intencionalidad, nos ayudarían a explicar exactamente cómo los procesos neurobiológicos en el cerebro causan nuestros estados subjetivos de advertir y sentir; cómo exactamente esos estados son realizados en las estructuras cerebrales; cómo exactamente funciona la conciencia en la economía global del cerebro y, por lo tanto, cómo funciona en nuestras vidas en general (Searle, 1997, p. 72). Clarificando así, nuestros conceptos de conciencia e intencionalidad.

Pues las respuestas a todas estas cuestiones causales, nos permitirían reformular y redefinir filosófica y neurocientificamente dichos conceptos de una manera que puedan ser epistemológicamente accesibles a partir de criterios objetivos. La solución a todas estas dificultades, nos arrojaría una luz sobre la manera en que debemos entender nuestros conceptos, puesto que la comprensión de la conciencia y los procesos neurobiológicos que la sustentan biológicamente, implica también la comprensión gramatical y conceptual del fenómeno conciencia.

De esta manera, nos despojaríamos del problema que ha perseguido como sombra a la filosofía del siglo XX cada vez que se ha ocupado de la subjetividad: el problema de las otras mentes. Es decir, la clarificación y la redefinición filosófica y neurocientifica de los conceptos de conciencia e intencionalidad, nos ayudarían a rediseñar el argumento de la habitación china, de un modo que no excluya al resto de los seres humanos de la categoría de sujetos mentales intencionales y, por consiguiente, le

daría a nuestros concepto criterios objetivos sobre los cuales puedan ser evaluados, comprendidos y accesibles.

Esto, por el hecho, de que el estudio neurocientifico de la conciencia se realiza sobre la base de un marco conceptual que se va redefiniendo filosóficamente a medida que ampliamos objetivamente nuestros conocimientos acerca de los misterios que la envuelven. Así, vemos que el estudio de la conciencia como un fenómeno biológico como cualquier otro, tal como nos lo propone Searle desde el enfoque neurobiológico, nos ayudaría a futuro, comprender los misterios de la conciencia y rediseñar objetivamente el concepto de la misma. De esta manera, evitaríamos reducir nuestros conceptos a criterios subjetivos, que no hacen otra cosa que conducirnos al solipsismo.

Bibliografía

- Albarán, E. S. (2004). *El experimento mental de la habitación china: Máquinas entre la semántica y la sintaxis.* Bogotá: Editor Fundación Método.
- Arbelo, L. A. (2001). *¿Tiene ventanas la habitación china? La imposibilidad del supuesto de la sustitución en Searle*. *Laguna*, 79-94.
- Descartes, R. (1977). *Meditaciones Metafísicas.* Madrid: Alfaguara.
- Descartes, R. (1982). *Discurso del método. .* Madrid: Espasa-Calpe.
- Descartes, R. (2009). *Las Pasiones del Alma.* México: Coyoacan.
- Moural, J. (2003). The chinese room argument. En B. Smith, *John Searle* (págs. 214-260). New York: Cambridge University Press.
- Navarro, J. (2005). C*ómo salir de la habitacion china: Conciencia e intencionalidad en las otras mentes*. *THÉMATA*, 267-275.
- Penrose, R. (1996). *La mente nueva del emperador.* Ciudad de México: CONSEJO NACIONAL DE CIENCIA Y TECNOLOGÍA, FORNDO DE CULTURA ECNÓMICA. MÉXICO.
- Roldán, M. C. (2001). *El descubrimiento de la mente: de san Agustín a Descartes*. *ALMAMATER*, 41-46.
- Searle, J. (1985). *Mente, Cerebros y Ciencia.* Madrid: Ediciones Cátedra, S.A.
- Searle, J. (1990). ¿Es la mente un programa informático? *Investigación y Ciencia*, 9-16.
- Searle, J. (1992). *Intencionalidad.* Madrid: Técnos.
- Searle, J. (1994). Mentes, cerebros y programas . En M. A. Boden, *Filosofía de la inteligencia artificial* (págs. 82-104). Ciudad de México: MÉXICO: FONDO DE CULTURA ECONÓMICA.
- Searle, J. (1995). Mentes y cerebros sin programas. En M. Boden, *Filosofía de la mente y ciencia cognitiva* (págs. 413-443). Barcelona: Paidós.
- Searle, J. (1996). *El redescubrimiento de la mente.* Barcelona: Crítica, Grijalbo Mondadori.
- Searle, J. (1997). *El misterio de la conciencia.* Barcelona : Paídos.
- Searle, J. (1997). *La construcción de la realidad social.* Barcelona: Paidós.

- Searle, J. (12 de Junio de 2007). *La conciencia.* Obtenido de Diálogos de Bioética: www.dialogos.unam.mx
- Turing, A. (1994). La maquinaria de computación y la inteligencia. En M. A. Boden, *Filosofía de la inteligencia artificial* (págs. 53-81). Ciudad de México: MÉXICO: FONDO DE CULTURA ECONÓMICA.